

## THESE

Présentée à

**L'UNIVERSITE DE BORDEAUX 1**

Ecole Doctorale Sciences et Environnements

Par **François HUBERT**

Pour l'obtention du grade de

**DOCTEUR**

SPECIALITE: Ecologie évolutive, fonctionnelle et des communautés

# RECONSTRUCTIONS PHYLOGENETIQUES DU GENRE *QUERCUS* A PARTIR DE SEQUENCES DU GENOME NUCLEAIRE ET CHLOROPLASTIQUE

Soutenue le 21 juin 2013

Devant la commission d'examen formée de :

Mme Nathalie FRASCARIA-LACOSTE	Professeur AgroParisTech, Orsay	Rapporteur
Mr Daniel PRAT	Professeur de l'Université de Lyon 1	Rapporteur
Mr Richard MICHALET	Professeur de l'Université de Bordeaux 1	Rapporteur
Mr Jean-Pierre RENAUDIN	Professeur de l'Université de Bordeaux 1	Président
Mr Sylvain JEANDROZ	Professeur AgroSup, Dijon	Examineur
Mr Antoine KREMER	Directeur de Recherche, INRA, Bordeaux	Directeur de thèse

UMR BIOGECO 1202, 69 route d'Arcachon, 33612 CESTAS Cedex, France

N° ordre : 4804





---

# REMERCIEMENTS

---

---

Ces six années et demie passées au sein de l'INRA font suite à plusieurs passages au sein de divers laboratoires, nationaux ou internationaux. Je ne continuerai pas dans la voie de la recherche Française par la suite, l'expérience acquise étant un poids désormais puisque nous sommes devenus « un risque qu'il faut maîtriser collectivement » pour nos instances.

Je vais donc en profiter pour remercier toutes les personnes qui ont contribué de près ou de loin à ce que cela se déroule dans les meilleures conditions possibles en commençant du début de cette « expérience » jusqu'à maintenant.

Je remercie Hélène B. qui m'a donné ma chance en 2004 pour intégrer un laboratoire d'Eco-Toxicologie Aquatique, merci de m'avoir donné le goût du travail d'équipe et de m'avoir montré que le travail pouvait être encore plus qu'une passion !

Mon envie depuis l'entrée à l'Université étant de travailler dans un laboratoire de biologie moléculaire, j'ai pu avoir la chance de la réaliser au Québec en 2003-2004, et je remercie chaleureusement France D. ainsi que Sophie B. avec qui j'ai découvert les techniques couramment utilisées dans ce domaine.

Merci également à mes amies Québécoises Valérie, et Annie pour tous ces bons moments passés loin de la France, sans parler des « partys » et de ce que le Québec peut offrir aux « Maudits Français » !!!

De retour en France, j'ai rejoint le Sud où j'ai pu de nouveau travailler dans un laboratoire de biologie moléculaire, mais sur des acariens et non plus des mammifères. Merci Marie T. de m'avoir permis de rejoindre ton équipe, merci également à Sabine G.

Je remercierai uniquement Aude L. concernant cette année de DEA, 2004 n'étant pas un très bon cru, merci à toi mémère, en espérant encore pleins de repas suite aux dons du sang ...

Merci encore à toi Marie T. de m'avoir permis de rejoindre le CBGP pour mon premier CDD, après les acariens, j'ai ainsi fait connaissance avec les nématodes. Merci à Mireille F. ainsi qu'à Gwen G. pour votre encadrement !

Après presque une année passée en tant qu'équipier à MacDonald's, j'ai pu rejoindre en 2005 l'équipe de Pathologie Forestière de l'INRA de Villenave d'Ornon appartenant à l'unité Biogéco. Merci Cyril D. de m'avoir accepté et donné l'occasion de renouer avec la recherche. Je remercie également Olivier F. qui le mérite vraiment, n'est ce pas ?

Ce passage au sein de l'unité Biogéco ne s'est pas arrêté après ce stage, étant donné que j'ai par la suite continué dans le laboratoire où je suis depuis 2006 et qui fait également partie de cette unité.

Je souhaite à ce stade porter une attention particulière à Marie-Pierre R. pour sa gentillesse et sa bienveillance qui l'ont toujours caractérisée, ainsi que sa bonne humeur quotidienne !!!

Merci à vous Sophie G. et Olivier L. de m'avoir permis de travailler avec vous en tant « qu'esclave » pendant presque une année pour aider la thèse d'Olivier L. Merci Sophie G. pour ce magnifique costume de « reine des abeilles » que je n'ai pas encore eu l'occasion de porter ^^.

Henri C, bien que tu sois exilé maintenant, je tenais à te remercier pour ta présence au laboratoire ainsi que les pauses de 10 heures qui étaient bien souvent agrémentées de gâteaux, espérons que tu arrives toujours à maintenir cette tradition en Guyane !!!

Merci à l'équipe « Secrétariat » de l'unité pour votre bonne humeur et votre patience, merci donc à toi Véro L, mais également à vous Florence L et Chantal B.

Merci à Loic K., Jean-Marc F. et à Thierry L. pour votre aide lorsque l'informatique me jouait des tours, et vive Linux ;-).

Je remercie Christophe P. de m'avoir embauché quelques mois afin de faire un peu de terrain alors qu'il faisait beau, merci également à Pauline G.G et à Camille L pour m'avoir fait confiance pour le travail de labo et plus particulièrement à Pauline pour l'initiation au monde de l'analyse des séquences nucléotidiques !

Monde que je n'ai pas par la suite quitté, puisque j'ai eu la chance de travailler sur la plateforme de génotypage et de séquençage du site. Merci donc à toi Franck S. d'avoir su partager mon temps entre la plateforme et la suite du travail commencé avec Pauline G.G !

La mission qui m'a été confiée par la suite a sans aucun doute conditionné l'existence de ce projet de thèse. Merci à Antoine K. (qui est devenu mon directeur de thèse) ainsi qu'à Catherine B. de m'avoir fait confiance pour le séquençage du fameux fragment DT et des ses nombreux problèmes. Merci encore à Franck S. pour les nombreux tests que j'ai eu la liberté d'effectuer !!!

Avant de remercier les personnes ayant eu une influence particulière sur le déroulement de la thèse, je souhaite remercier mes collègues de bureau : Catherine B., Frédéric R. et Jean-Marc L. de m'avoir permis de toujours travailler dans de bonnes conditions.

Je remercie tout d'abord Alain F. de m'avoir fait confiance pour mener à bien (ou le mieux possible) la présente thèse. Merci pour les discussions que l'on a pu avoir à ce sujet, même si elles n'ont pas toujours été de tout repos ;-).

Merci Antoine K. d'avoir été mon directeur de thèse. Merci d'avoir su m'accorder du temps malgré ton emploi du temps plus que chargé ! Merci également de ne pas avoir baissé les bras comme la plupart lorsque nous avons dû nous battre pour que cette thèse puisse se finir dans un temps raisonnable !!!

Je remercie également tous les membres du jury pour le temps passé à lire et/ou écouter les conclusions de la présente étude. Il m'est agréable à ce stade de compter comme président du jury un de mes anciens enseignants de DEUG de l'Université de Bordeaux I.

Merci encore à toi Pauline G.G d'avoir rendu accessible le travail que tu as fait en amont, sans cela, nous n'aurions pas eu ou peu de résultats sur le génome nucléaire ....

Je remercie également Nicolas F. qui m'a beaucoup apporté au début de ce travail de thèse.

Merci Virgil.F de m'avoir fait confiance pour donner un cours magistral et des TDs dans les bâtiments de l'université ou j'ai moi-même été étudiant !

Ce travail de thèse a pris un côté « international » puisque j'ai eu la chance de pouvoir aller une semaine en Italie dans le laboratoire de Beppe V. et je l'en remercie chaleureusement.

De façon plus « locale », je remercie chaleureusement Vincent B. pour son aide sur l'univers des SuperArbres, mais surtout, Emmanuelle J. pour ses encouragements et son aide très précieuse lorsque le monde de la phylogénie me paraissait obscur !!!

Merci à toi Paul F. pour ces magnifiques cartes de répartitions, mais également de m'avoir poussé vers le monde du logiciel libre ... Je t'ai promis un paragraphe dans mes remerciements mais du coup tu en auras même deux !!!

Merci Cécile D pour les corrections orthographiques de dernière minute ;-)

De façon plus personnelle, je remercie tous les membres de l'équipe et particulièrement ceux avec qui j'ai partagé de bons moments. Merci à Benoit B., Tristan C. et Emilie C. pour les parties de ping-pong entre midi et deux ! Après votre départ, le ping-pong Pierrotonnais est resté vivant quelque temps, et je remercie donc à ce stade Paul F. et Cécile D. qui y ont participé (bien que le niveau était moins élevé ;-)) !

Merci à toi Emilie C., dommage que tu ais quitté Pierroton si rapidement, la politique ayant rendu ton départ inéluctable....

Merci Erwan pour ces longues balades dans le superbe domaine qu'est Pierroton ainsi que la recherche de champignons !

Les repas n'auraient pas été les mêmes sans nos discussions « philosophiques » et les « pauses » entre les plats. Merci donc à Christophe B. et à Lélia L. pour cela !!!!

Pour terminer, je souhaite remercier tous mes amis « proches » et « moins proches » qui se reconnaîtront mais plus particulièrement Paul F. et Cécile D. pour toutes les discussions que l'on a pu avoir au sujet de mes doutes et de mes sempiternelles questions, des activités sportives plus ou moins longues que l'on a eues (comme le foot US) mais également pour votre Humanité en général !!! J'espère que nos exploits à Sing Star et Just Dance se poursuivront ;-)) ...

## SOMMAIRE

---

---

<b>BIBLIOGRAPHIE.....</b>	<b>19</b>
<b>A1 TAXONOMIE.....</b>	<b>23</b>
<b>A1.1 CHEZ LES FAGACEES .....</b>	<b>23</b>
<b>A1.2 CHEZ LE GENRE QUERCUS.....</b>	<b>27</b>
<b>A2 PHYLOGENIES BASEES SUR DES CARACTERES MORPHOLOGIQUES .....</b>	<b>33</b>
<b>A2.1 RECONSTRUCTIONS PHYLOGENETIQUES INTERGENRES .....</b>	<b>35</b>
<b>A2.2 RECONSTRUCTIONS PHYLOGENETIQUES A L'INTERIEUR DU GENRE QUERCUS.....</b>	<b>37</b>
<b>A3 PHYLOGENIES BASEES SUR DES CARACTERES MOLECULAIRES .....</b>	<b>39</b>
<b>A3.1 PHYLOGENIES BASEES SUR DES MARQUEURS CHLOROPLASTIQUES .....</b>	<b>41</b>
A3.1.1 CHEZ LES FAGACEES.....	41
A3.1.2 CHEZ LE GENRE QUERCUS.....	41
<b>A3.2 PHYLOGENIES BASEES SUR DES MARQUEURS NUCLEAIRES .....</b>	<b>45</b>
A3.2.1 CHEZ LES FAGACEES .....	47
A3.2.2 CHEZ LE GENRE <i>QUERCUS</i> .....	49
<b>A3.3 REFERENCES BIBLIOGRAPHIQUES .....</b>	<b>69</b>

**A4 BIOGEOGRAPHIE HISTORIQUE DU GENRE QUERCUS ..... 73**

**A4.1 EVOLUTION DU GENRE QUERCUS ..... 73**

A4.1.1 DONNEES FOSSILES ..... 73

A4.1.2 SCÉNARIOS ÉVOLUTIFS PRINCIPAUX ÉVOQUÉS ..... 77

**A4.2 APPORTS DE LA BASE DE DONNEES FOSSIL MAP A LA CALIBRATION ..... 83**

A4.2.1 DATATION DU GROUPE CYCLOBALANOPSIS (TABLEAU A7) : ..... 87

A4.2.2 DATATION DU GROUPE CERRIS (TABLEAU A8) : ..... 89

A4.2.3 DATATION DU GROUPE ILEX (TABLEAU A9) : ..... 89

A4.2.4 DATATION DU GROUPE QUERCUS S.S (TABLEAU A10) : ..... 89

A4.2.5 DATATION DU GROUPE PROTOBALANUS (TABLEAU A11) : ..... 91

A4.2.6 DATATION DU GROUPE LOBATAE (TABLEAU A12) : ..... 91

A4.2.7 COMPARAISON ENTRE LES DEUX TYPES DE DATATIONS (TABLEAU A13) : ..... 91

**A4.3 REFERENCES BIBLIOGRAPHIQUES ..... 95**

**A5 CONTEXTE, OBJECTIF ET METHODOLOGIE GENERALE DU PRESENT TRAVAIL DE THESE ..... 97**

**B1 MATERIEL ET METHODES MOLECULAIRES : .....103**

**B1.1 ECHANTILLONNAGE .....103**

<b>B1.2 EXTRACTION D'ADN, AMPLIFICATION ET SEQUENÇAGE .....</b>	<b>105</b>
<b>B1.3 RECONSTRUCTIONS PHYLOGENETIQUES.....</b>	<b>109</b>
<b><u>B2 CONSERVED ORTHOLOG SET MARKERS (COS) .....</u></b>	<b><u>115</u></b>
<b>B2.1 OBJECTIF .....</b>	<b>115</b>
<b>B2.2 CHOIX DES MARQUEURS.....</b>	<b>117</b>
<b>B2.3 RESULTATS GENE PAR GENE .....</b>	<b>123</b>
DESCRIPTION DU COS A6785 (FIGURE B3).....	123
DESCRIPTION DU COS A5799 (FIGURE B4).....	123
DESCRIPTION DU COS H10808 (FIGURE B5).....	125
DESCRIPTION DU COS H9687 (FIGURE B6).....	125
<b>B2.4 RESULTAT SUPER ARBRE (FIGURE B7).....</b>	<b>127</b>
<b>B2.5 RESULTAT SUPER MATRICE (FIGURE B8) .....</b>	<b>127</b>
<b>B2.6 DISCUSSION/CONCLUSION .....</b>	<b>131</b>
<b>B2.7 DEROULEMENT DE CETTE PARTIE DE L'ETUDE .....</b>	<b>133</b>
<b>B2.8 RÉFÉRENCES BIBLIOGRAPHIQUES .....</b>	<b>135</b>
<b><u>B3 PHYLOGENETIC RELATIONSHIPS BETWEEN INFRAGENERIC GROUPS IN THE</u></b>	
<b><u>GENUS QUERCUS .....</u></b>	<b><u>137</u></b>

<b>B3.1 INTRODDUCTION .....</b>	<b>141</b>
<b>B3.2 MATERIAL AND METHODS .....</b>	<b>145</b>
B3.2.1 SPECIES SAMPLING.....	145
B3.2.2 DNA EXTRACTION, MARKER SELECTION, AMPLIFICATION AND SEQUENCING.....	145
B3.2.3 PHYLOGENETIC RECONSTRUCTIONS.....	149
<b>B3.3 RESULTS .....</b>	<b>153</b>
B3.3.1 PCR AND SEQUENCING RESULTS.....	153
B3.3.2 SINGLE GENE ANALYSIS.....	153
B3.3.3 COMBINED ANALYSES .....	155
<b>B3.4 DISCUSSION .....</b>	<b>159</b>
B3.4.1 TREE TOPOLOGIES AND OUTGROUP CHOICE .....	159
B3.4.2 SUBGENERIC AND INFRAGENERIC SUBDIVISION OF GENUS QUERCUS.....	161
B3.4.3 POORLY SUPPORTED PHYLOGENETIC RESOLUTION WITHIN THE GENUS QUERCUS IN DEEPER NODES .....	165
<b>B3.5 BILAN DES RECONSTRUCTIONS PHYLOGENETIQUES BASES SUR DES GENES NUCLEAIRES.....</b>	<b>167</b>
<b>B3.6 PHYLOGENIE CALIBREE BASEE SUR SIX MARQUEURS NUCLEAIRES .....</b>	<b>171</b>
<b>B3.7 DEROULEMENT DE CETTE ETUDE .....</b>	<b>179</b>
<b>B3.8 REFERENCES .....</b>	<b>181</b>

<b><u>B4 ETUDE DU GENOME CHLOROPLASTIQUE .....</u></b>	<b><u>189</u></b>
<b>B4.1 INTRODUCTION .....</b>	<b>191</b>
<b>B4.2 MATERIEL ET METHODES .....</b>	<b>199</b>
B4.2.1 ECHANTILLONNAGE .....	199
B4.2.2 AMPLIFICATION DE L'ADN CHLOROPLASTIQUE ET SEQUENCAGE.....	203
B4.2.3 ASSEMBLAGE DU GENOME ET ANNOTATION.....	209
B4.2.4 RECONSTRUCTIONS PHYLOGENETIQUES .....	209
<b>B4.3 RESULTATS .....</b>	<b>219</b>
B4.3.1 RESULTAT SUPERMATRICE SUR LE GENOME TOTAL.....	219
B4.3.2 RESULTATS SUPERMATRICE SUR LE SOUS ENSEMBLE .....	221
B4.3.3 RESULTAT SUPERARBRE SUR LE SOUS ENSEMBLE .....	221
<b>B4.4 DISCUSSION/CONCLUSION .....</b>	<b>223</b>
<b>B4.5 DEROULEMENT DE CETTE ETUDE .....</b>	<b>231</b>
<b>B4.6 RÉFÉRENCES BIBLIOGRAPHIQUES .....</b>	<b>233</b>
<b><u>B5 COMPARAISON ENTRE LE GENOME NUCLEAIRE ET LE GENOME</u></b>	
<b><u>CHLOROPLASTIQUE : APPORTS A LA BIOGEOGRAPHIE HISTORIQUE DU GENRE</u></b>	
<b><u>QUERCUS .....</u></b>	<b><u>239</u></b>
<b><u>B6 CONCLUSION GENERALE - PERSPECTIVES .....</u></b>	<b><u>245</u></b>

**B6.1 RÉFÉRENCES BIBLIOGRAPHIQUES .....255**

**B7 : ANNEXES .....257**

---

INTRODUCTION GENERALE ET  
PRESENTATION DU PROJET  
PHYLOSPACE

---



Depuis l'apparition de la vie sur la planète, le nombre d'espèces ainsi que leur diversité n'a cessé d'augmenter au cours du temps. Cette augmentation n'a pas été régulière, mais a été caractérisée d'après les données géologiques par des taux élevés de spéciation suivis par de longues périodes de stases. Cinq périodes d'extinctions massives ont été dénombrées dans les archives géologiques dont la plus importante est datée de -250 millions d'années. Plus de 50% des espèces animales se seraient éteintes, la perte dans les océans aurait même atteint les 96% !!! Cette perte de diversité spécifique serait due à de profonds changements climatiques, changements certainement provoqués par de vastes éruptions volcaniques, des collisions d'astéroïdes, la libération massive de méthane par le plancher océanique, des orogénèses ... Il a fallu par exemple plus de 50 millions d'années pour retrouver la biodiversité perdue suite à la crise Permienne.

Les taux d'extinction actuels sont considérés par les spécialistes comme étant de 100 à 1000 fois plus importants que le taux « naturel », et plus de 99% de celles-ci sont le fait des activités anthropiques (Lovei, 2001; Mace, 2005; Pimm and Jenkins, 2005). Cet épisode de bouleversement de la variabilité spécifique est considéré comme la sixième période d'extinction massive (Barnosky *et al.*, 2011). C'est dans ce cadre qu'un appel à projets ANR (Agence Nationale pour la Recherche) a été déposé en 2010 dont l'intitulé est : « La 6ème extinction : quantifier la perte de diversité biologique ; comprendre et agir sur les processus biologiques, économiques et sociaux qui l'accompagnent ».

Les objectifs du programme "La 6ème extinction" sont de faire progresser les connaissances sur les facteurs pouvant conduire à une perte de biodiversité et sur les conséquences de cette perte à la fois sur le fonctionnement des écosystèmes et sur les développements socio-économiques. L'ensemble de ces données pourra constituer une base dans le cadre de l'élaboration de stratégies d'orientation pour un développement durable des activités humaines.

L'appel à projets s'articule autour de trois axes :

1. Documenter et caractériser l'érosion de la biodiversité et la dégradation des services écosystémiques.
2. Scénariser et modéliser les changements de la biodiversité.
3. Comprendre les interactions dynamiques entre les processus économiques et sociaux et la réduction de la biodiversité.

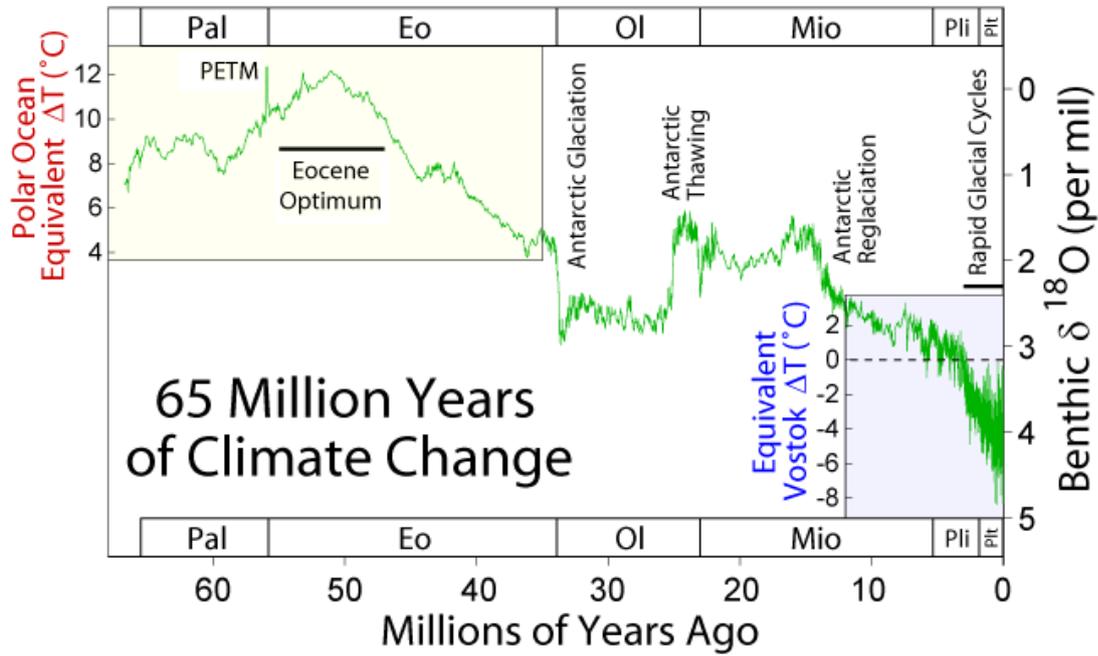


FIGURE 11: EVOLUTION DU CLIMAT DURANT LES 60 DERNIERS MILLIONS D'ANNEES. LE PETM CORRESPOND AU MAXIMUM THERMIQUE RELEVE A LA TRANSITION PALEOCENE-EOCENE DURANT UNE TRES COURTE PERIODE DE TEMPS. SOURCE : [HTTPS://EN.WIKIPEDIA.ORG/WIKI/PALEOCENE%E2%80%93EOCENE](https://en.wikipedia.org/wiki/Paleocene%E2%80%93Eocene)

Le projet global auquel se rattache la présente thèse s'intitule « PhyloSPACE : intégrer des cophylogénies, des changements d'aires et des cartes paléoclimatiques pour inférer l'histoire d'associations sous changements climatiques ». Ce projet, coordonné par Alain Franc, regroupe plusieurs équipes travaillant sur différents modèles biologiques dont le chêne (*Quercus sp.*). L'objectif de ce projet est d'étudier l'évolution du nombre d'espèces dans le temps, et de mettre ceci en lien avec les données climatiques passées à l'aide de cartes paléoclimatiques. Le réchauffement climatique actuel très rapide est comparable à ce qu'il s'est passé il y a plus de 55 millions d'années et que l'on appelle le Maximum Thermique Paléocène-Eocène ou PETM (Figure I1) : une augmentation rapide de la température du globe.

L'esprit de cette thèse s'inscrit donc dans la volonté d'obtenir un arbre phylogénétique du chêne décrivant l'évolution dans le temps de ce genre -idéalement calibré avec des données fossiles- afin de permettre de voir si ce maximum thermique a eu des répercussions sur la diversité passée spécifique du chêne. La description de ce phénomène passé permettant de prédire les conséquences du réchauffement climatique actuel.



## BIBLIOGRAPHIE

---

- Barnosky, A. D., *et al.* (2011). "Has the Earth's sixth mass extinction already arrived?" Nature **471**(7336): 51-57.
- Lovei, G. (2001). Extinctions, modern examples of. Encyclopedia of Biodiversity. San Diego, Academic Press. **2**: 731-744.
- Mace, G. M. (2005). Biodiversity. Millennium Ecosystem Assessment (MEA), Ecosystems and Human Well-Being : Current State and Trends. Washington D. C., Island Press. **1**: 77-122.
- Pimm, S. L. and C. N. Jenkins (2005). "Sustaining the Variety of Life." Scientific American **293**(3): 66-73.



---

PARTIE A

INTRODUCTION BIBLIOGRAPHIQUE

---

TABLEAU A1 : DISTRIBUTION GEOGRAPHIQUE ET NOMBRE D'ESPECES CHEZ LES DIFFERENTS GENRES DES FAGACEES (ZHENG, 1985; JONES, 1986; XU, 2004; MANOS *ET AL.*, 2008).

<b>Genre</b>	<b>Nombre d'espèces</b>	<b>Répartition Géographique</b>
<i>Castanea</i>	12	Europe, Asie, Amérique du Nord
<i>Castanopsis</i>	100-200	Sud Est Asiatique
<i>Chrysolepis</i>	2	Amérique du Nord (côte Pacifique)
<i>Fagus</i>	13	Europe, Asie, Amérique du Nord (une espèce)
<i>Lithocarpus</i>	300	Asie
<i>Notholithocarpus</i>	1	Amérique du Nord
<i>Quercus s.l</i>	<ul style="list-style-type: none"> <li>• 300-450 (<i>Quercus s.s</i>)</li> <li>• 150 (<i>Cyclobalanopsis</i>)</li> </ul>	<ul style="list-style-type: none"> <li>• Europe, Asie, Amérique du Nord, Centrale et du Sud, Afrique du Nord</li> <li>• Sud de la Chine, Indochine</li> </ul>
<i>Trigonobalanus</i>	3	Asie, Amérique du Sud (Colombie)

---

## A1 TAXONOMIE

---

---

### A1.1 CHEZ LES FAGACEES

---

La famille des Fagacées qui appartient au groupe des Angiospermes et à l'ordre des Fagales est proche génétiquement des Cucurbitales, des Rosales, puis des Fabales (Figure A1). Cette famille comprends plus de 1000 espèces sur le globe, et regroupe 8 genres actuellement : *Fagus*, *Castanea*, *Castanopsis*, *Chrysolepis*, *Trigonobalanus*, *Lithocarpus*, *Notholithocarpus* et *Quercus* (Tableau A1).

Le nombre de sous familles composant les Fagacées est compris entre 2 et 4 (Tableau A2), suivant que l'on ajoute aux Fagoideae et aux Castaneoideae les sous familles des Quercoideae et des Trigonobalanoideae.

La composition de la famille des Castaneoideae ne varie pas entre les différents types de classification et regroupe les genres *Castanea*, *Castanopsis*, *Lithocarpus* et *Chrysolepis* (Tableau A2). La classification de Nixon et Crepet (1989) qui subdivise la famille des Fagaceae en deux sous familles, se base sur les différents types de pollinisation que l'on retrouve dans cette famille : le syndrome anémophile souvent associé à une germination épigée (à l'exception de *Quercus*, dont la position des glands est hypogée) qui regroupe les genres *Quercus*, *Fagus*, *Trigonobalanus*, et le syndrome entomophile associé à une germination hypogée et à des cupules de type dichasium qui regroupe les autres genres (*Castanea*, *Castanopsis*, *Lithocarpus* et *Chrysolepis*).

Il est intéressant de noter la place particulière qu'occupent les Fagacées dans l'ordre des Fagales car les autres familles ont une pollinisation uniquement anémophile.

La subdivision de la famille en genres, ne fait pas également l'unanimité au sein des botanistes : le genre *Notholithocarpus* était jusqu'en 2008 (Manos *et al.*, 2008) associé au genre *Lithocarpus* sur la base de leur ressemblance morphologique (Hooker and Arnott, 1840). Tous les membres du genre *Lithocarpus* se trouvent sur le continent asiatique à l'exception d'une seule espèce : *Lithocarpus densiflorus* qui se retrouvait sur le sol américain. Manos (Manos, 2008), à proposé d'assigner cette espèce au rang de genre, d'où le nom de *Notholithocarpus*.

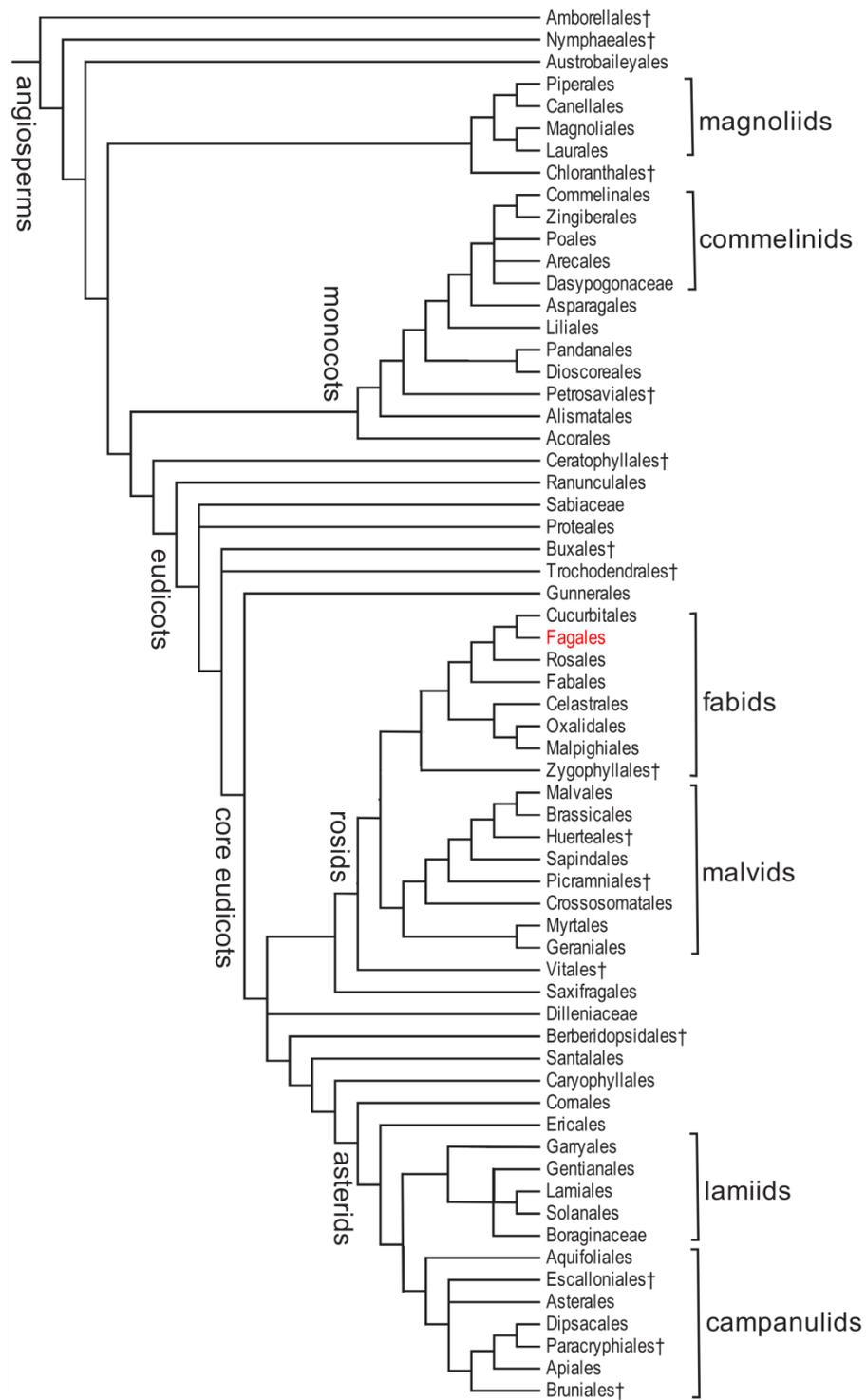


FIGURE A1 : ARBRE PHYLOGENETIQUE REPRESENTANT LES ORDRES ET QUELQUES FAMILLES D'ANGIOSPERMES, (BREMER *ET AL.*, 2009).

Le genre *Trigonobalanus* au sens large, regroupe selon Nixon (Nixon and Crepet, 1989) le genre *Trigonobalanus* au sens strict, le genre *Formanodendron* et le genre *Colombobalanus* qui sont composés chacun par une seule espèce.

Le genre *Quercus* au sens large, est scindé d'après Camus (Camus, 1936-1954) en deux sous-genres : le sous-genre *Euquercus* (sensu Camus) et le sous-genre *Cyclobalanopsis*.

Le nombre d'espèces composant les 8 genres précédemment définis est inégal : les genres *Quercus*, *Lithocarpus* et *Castanopsis* sont constitués de plusieurs centaines d'espèces alors que les autres comportent quelques dizaines voire une espèce par genre (Tableau A1).

Les Fagacées sont distribuées dans la majeure partie de l'hémisphère nord et dominant la flore actuelle (Camus, 1936-1954) avec un nombre de genres plus élevés en Asie du sud est, suggérant ainsi un probable centre de diversité.

Les différentes espèces sont distribuées dans toutes les régions tempérées, dans les régions tropicales montagneuses de l'Asie du Sud, de l'Amérique Centrale et du Sud et les régions méditerranéennes de l'Afrique et de l'Europe. Il n'y a pas de Fagacées dans la partie centrale de l'Afrique du Nord, la sécheresse du climat aurait été un facteur limitant, mais les Fagacées y auraient été présents dans le passé (Raven and Axelrod, 1974).

les genres *Quercus*, *Fagus* et *Castanea* se trouvent en Europe, en Asie et en Amérique, *Trigonobalanus* en Asie et Amérique du Sud, *Lithocarpus* et *Castanopsis* se retrouvent uniquement en Asie, *Notholithocarpus* et *Chrysolepis* uniquement en Amérique du Nord (Tableau A1 et Figure A2).

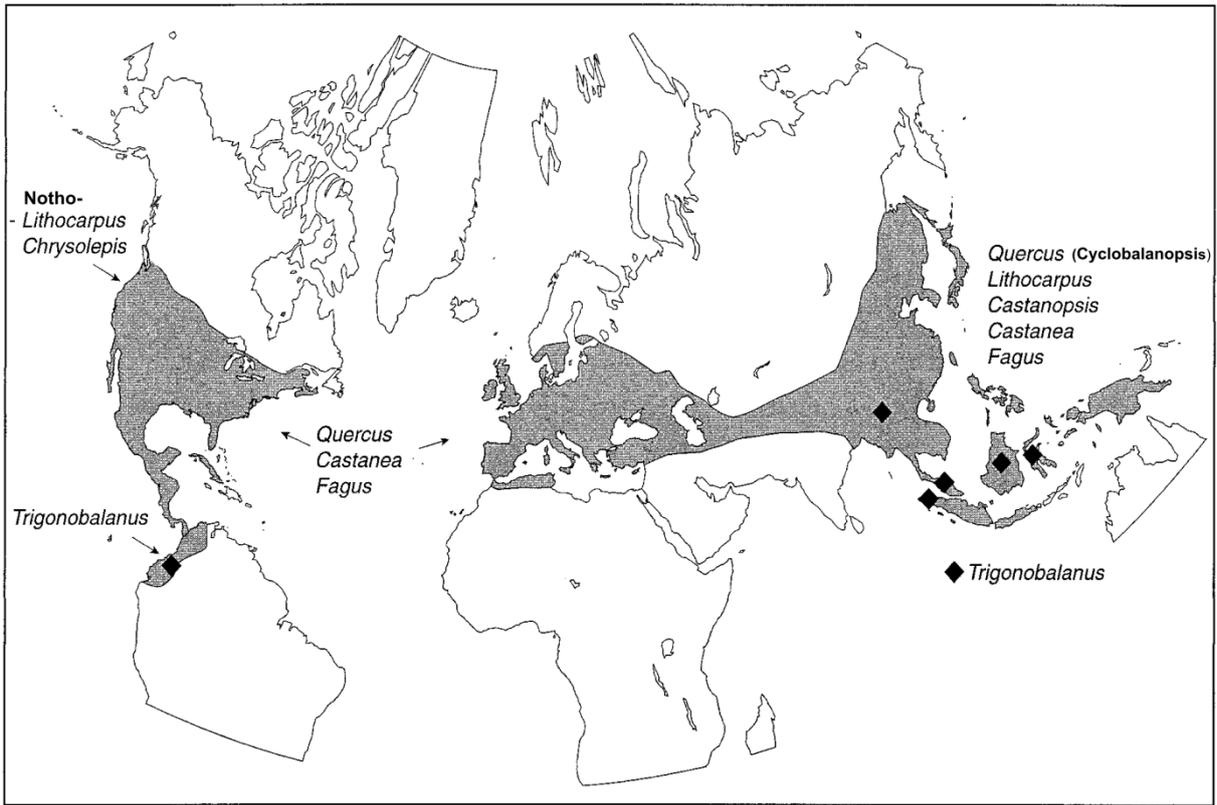


FIGURE A2 = REPARTITION GEOGRAPHIQUE DES 8 GENRES, TIRÉE DE MANOS AND M. STANFORD 2001, ET D'APRES SOEPADMO, 1972; NIXON AND CREPET, 1989; NIXON, 1997).

## A1.2 CHEZ LE GENRE QUERCUS

---

Le genre *Quercus*, selon Camus inclue les sous-genres *Euquercus* et *Cyclobalanopsis*, comprend entre 300 et 600 espèces selon les auteurs. Ce genre se retrouve en Asie, en Amérique du Nord, en Europe et en Afrique (Axelrod, 1983).

Le sous-genre *Cyclobalanopsis* se trouve uniquement en Asie du Sud Est et comporte 76 espèces, le sous-genre *Euquercus* comporte 49 espèces en Asie, 234 au Nouveau Monde et 22 en Europe (Figure A3).

La monographie botanique la plus complète du genre *Quercus*, incluant les espèces présentes sur l'ensemble des continents, est celle de Camus (*Camus, 1936-1954*) et divise le sous-genre *Euquercus* en six sections : la section des chênes blancs (*Lepidobalanus* ou *Quercus*), la section des chênes rouges (*Erythrobalanus* ou *Lobatae*), les chênes intermédiaires (*Protobalanus*), les *Mesobalanus*, les *Macrobalanus* et les *Cerris*. Ces différentes sections seront discutées ultérieurement dans le manuscrit, en effet, divers amendements nous semblent nécessaires afin de refléter au mieux la diversité trouvée au sein de ce genre comme l'inclusion des *Mesobalanus* et des *Macrobalanus* au sein des *Lepidobalanus*.

Les caractères permettant de diviser ce genre en sections sont :

- \* La forme du style (axe supportant le stigmate, dans lequel se développe le tube pollinique lors de la fécondation) qui peut être de longueur et d'épaisseur différente ;

- \* La localisation des ovules qui peuvent se situer sur la partie inférieure, latérale ou supérieure du fruit ;

- \* L'endocarpe (paroi du fruit) qui peut être ou non duveteux ;

- \* La symétrie ou la non symétrie des cotylédons.

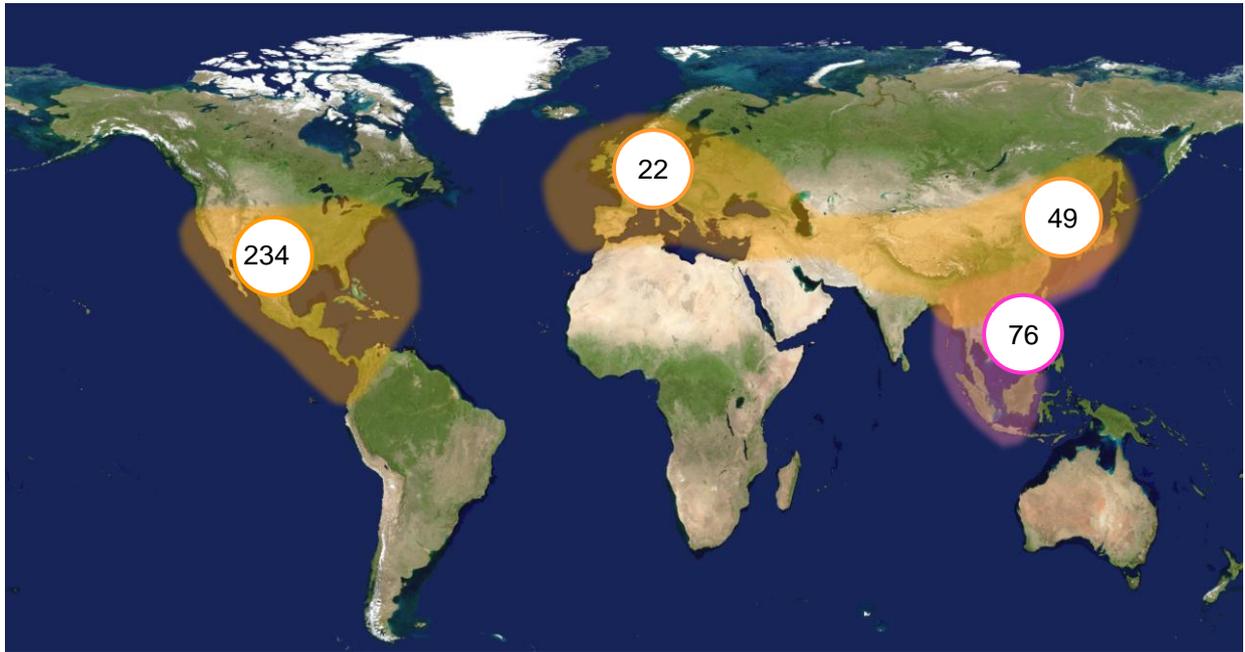


FIGURE A3 : REPARTITION GEOGRAPHIQUE ACTUELLE DU GENRE *QUERCUS* ET NOMBRE D'ESPECES (SCHWARZ *ET AL.*, 1964; NIXON, 1993; VALENCIA, 2004; MENITSKY, 2005). EN JAUNE, LE SOUS-GENRE EUQUERCUS (SENSU CAMUS) ET EN ROSE LE SOUS-GENRE CYCLOBALANOPSIS.

En s'appuyant sur la classification de Camus (Camus, 1936-1954), la section des *Lepidobalanus* se retrouve en Europe, en Asie, et en Amérique avec un nombre d'espèces variant suivant les auteurs entre 167 et 172 (Tableau A3), les *Erythrobalanus* sont uniquement inféodés à l'Amérique avec un total compris entre 90 et 111 espèces (Tableau A3), tout comme les *Protobalanus* qui sont composés par neuf espèces (Tableau A3), alors que la section *Cerris* se retrouve uniquement en Asie avec 29 espèces (Tableau A3).

En Amérique, la plus grande richesse spécifique chez les *Erythrobalanus* et les *Lepidobalanus* se trouve au Mexique, avec un nombre d'espèces compris entre 55 et 76 pour les *Erythrobalanus* et 76 à 81 pour les *Lepidobalanus* (Tableau A3), suggérant un très probable centre de diversité.

En Eurasie, la diversité spécifique la plus élevée se trouve en Asie, totalisant 25 espèces de *Lepidobalanus* sur les 42 présentes, et 24 sur 29 espèces de la section *Cerris* (Tableau A3).

Ces rappels de botanique soulèvent des problèmes de nomenclature et de classification. Ces deux éléments ont continuellement évolué au cours du temps selon les auteurs, depuis que le genre *Quercus* a été décrit pour la première fois par Ørsted en 1871 (Ørsted, 1867), (voir la revue sur le sujet dans Xu, 2004 et Denk and Grimm, 2010).

De nombreuses classifications ont par ailleurs été proposées depuis cette date, attribuant différent rangs taxonomiques (sous-genre, section...) à des groupes d'espèces. Leur énumération alourdirait et compliquerait la lecture de ce texte. Nous avons au début de ce texte privilégié la nomenclature et la classification de Camus (1936-1954), en raison du fait qu'il s'agissait encore de l'unique monographie mondiale du genre. Cette nomenclature a cependant été revue par Nixon (1993), au vu des critères utilisés en la matière (notamment en termes d'antériorité des noms) et cette révision est aujourd'hui communément admise.

TABLEAU A2 : DIFFERENTES SUBDIVISIONS DE LA FAMILLE DES FAGACEES (FORMAN, 1966; SOEPADMO, 1972; JONES, 1986; NIXON AND CREPET, 1989; OHBA, 1989; XU, 2004), D'APRES XU (2004).

<b>Sous Famille</b>	<b>Ohba</b>	<b>Nixon et Crepet</b>	<b>Jones</b>	<b>Forman, Soepadmo</b>
Fagoideae	<i>Fagus</i>	<i>Fagus</i> <i>Quercus</i> <i>Trigonobalanus</i>	<i>Fagus</i>	<i>Fagus</i>
Castaneoideae	<i>Castanea</i> <i>Castanopsis</i> <i>Lithocarpus</i>	<i>Castanea</i> <i>Castanopsis</i> <i>Lithocarpus</i> <i>Chrysolepis</i>	<i>Castanea</i> <i>Castanopsis</i> <i>Lithocarpus</i> <i>Chrysolepis</i>	<i>Castanea</i> <i>Castanopsis</i> <i>Lithocarpus</i>
Quercoideae	<i>Quercus</i>		<i>Quercus</i>	<i>Quercus</i> <i>Trigonobalanus</i>
Trigonobalanoideae	<i>Trigonobalanus</i>			

Par ailleurs, et ceci nous oblige à anticiper sur les reconstructions phylogénétiques qui vont suivre dans le texte, la communauté scientifique travaillant sur la botanique, la taxonomie du genre *Quercus* s'accorde aujourd'hui sur deux faits :

1) La subdivision du genre en deux sous-genres :

- Le sous-genre *Cyclobalanopsis* (selon Camus et Nixon) ou sous-genre *Cyclobalanoides* (au sens Menitsky) ;
- Le sous-genre *Euquercus* (selon Camus), ou sous-genre *Quercus* (sensu Menitsky) ou *Quercus sensu lato* (selon Nixon), nom que nous utiliserons désormais.

2) La subdivision du sous-genre *Quercus s.l.* en cinq groupes infragénériques, que nous désignerons par groupe :

- Le groupe *Quercus sensu stricto* (*Quercus s.s.*), ou chênes blancs ;
- Le groupe *Lobatae*, ou chênes rouges ;
- Le groupe *Protobalanus*, ou chênes intermédiaires ;
- Le groupe *Cerris*, ou chênes *cerris* ;
- Le groupe *Ilex*, ou chêne verts.

Au plan botanique, cette subdivision emprunte en fait des éléments issus de plusieurs classifications : celles de Camus (1936-1954), de Nixon (1993) et de Menitsky (1984, 2005). Les deux premières considérant l'ensemble du genre, alors que la seconde se limite aux chênes eurasiatiques. Les correspondances entre ces classifications sont données dans le tableau A4.

Ce tableau illustre cependant l'ambiguïté relative à l'utilisation du terme *Quercus*, qui peut soit désigner un genre, un sous-genre ou une section dans la nomenclature aujourd'hui utilisée. C'est la raison pour laquelle nous lui adjoignons le suffixe *s.l.* ou *s.s.* pour désigner le sous-genre ou la section.

TABLEAU A3 : REPARTITION GEOGRAPHIQUE DES DIFFERENTES SECTIONS DU GENRE *QUERCUS* (SCHWARZ ET AL., 1964; NIXON, 1993; VALENCIA, 2004; MENITSKY, 2005) EN SE BASANT SUR LA CLASSIFICATION DE CAMUS (CAMUS, 1936-1954).

		<b>Amérique du Nord (USA+Canada)</b>	<b>Mexique</b>	<b>Amérique centrale + du Sud</b>	<b>Europe + Afrique du Nord</b>	<b>Asie</b>
	<b><i>Euquercus</i></b>	87-90	135-161	30-36	22	49
<i>Lepidobalanus</i>		49	76-81	0	17	25
<i>Erythrobalanus</i>		34	55-76	1	0	0
<i>Protobalanus</i>		5	4	0	0	0
<i>Cerris</i>		0	0	0	5	24
	<b><i>Cyclobalanopsis</i></b>	0	0	0	0	76

## A2 PHYLOGENIES BASEES SUR DES CARACTERES MORPHOLOGIQUES

---

La phylogénie s'intéresse à la construction des relations de parenté entre taxons à partir d'observation actuelle, retraçant ainsi l'évolution d'un groupe d'espèces. A partir de données actuelles, ces méthodes sont capables d'inférer des scénarios évolutifs passés, schématisés sous forme d'arbre dont les nœuds représentent les espèces ancestrales dont dérivent les espèces actuelles.

Afin d'établir des relations phylogénétiques entre taxons à partir de données morphologiques, il est nécessaire de faire appel à des méthodes de cladistique. La méthode cladistique se base sur la polarité des caractères et peut être utilisée pour des caractères morphologiques ou moléculaires. En effet, un caractère peut être ancestral (ou plésiomorphe) ou dérivé (apomorphe). La cladistique reconnaît des groupes à partir du moment où ils sont formés par des caractères apomorphes = une évolution apportée au groupe. L'arbre phylogénétique choisi sera celui (ou ceux) qui sera le plus parcimonieux = celui qui demandera le moins de changements d'états au vue du ou des caractères étudiés.

Il existe plusieurs inconvénients à ce type d'analyse car il est nécessaire d'avoir des arguments pour pouvoir polariser les états de caractères (par exemple l'utilisation de fossiles : l'état ancestral sera défini par l'état présent dans les registres fossiles), et il n'y a pas de modèle d'évolution.

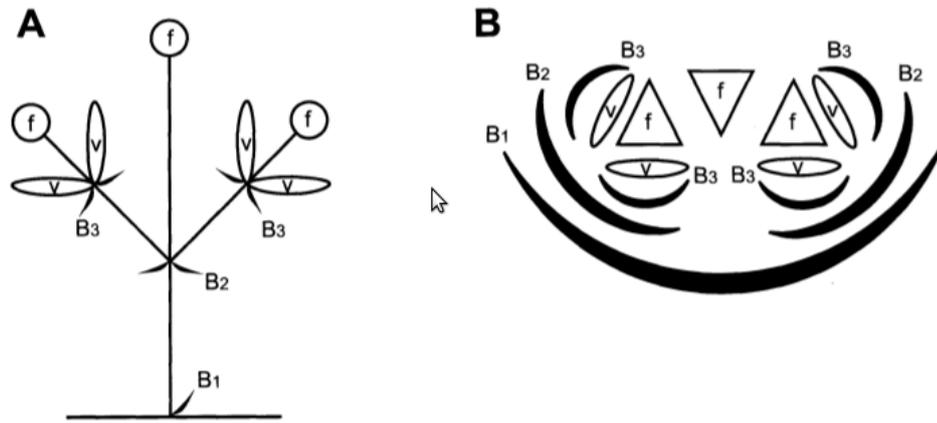
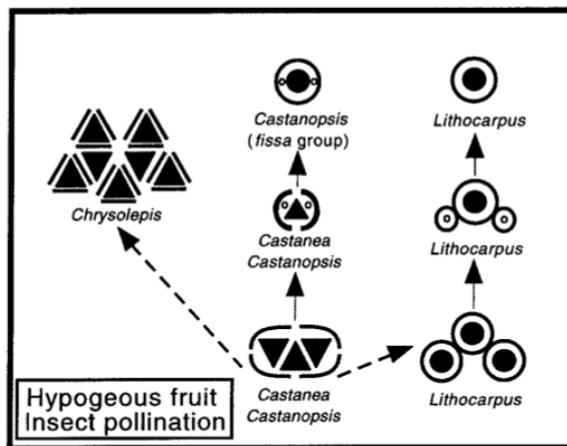


FIGURE A4 : DIAGRAMME D'UNE CUPULE DE TYPE DICHASIAL (NIXON AND CREPET, 1989; OH AND MANOS, 2008).

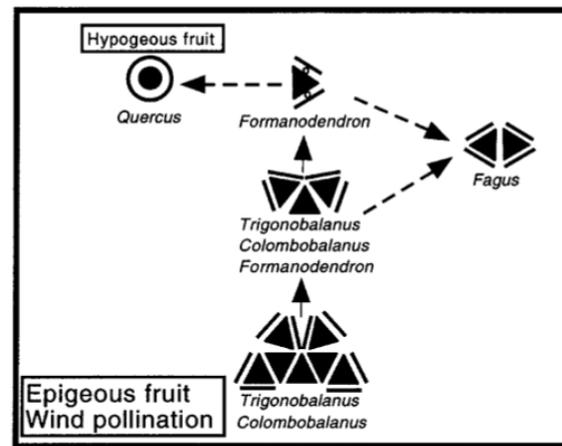
A : INFLORESCENCE HYPOTHETIQUE. CHAQUE VALVE REPRESENTE UNE BRANCHE STERILE.

B : DEVELOPPEMENT INITIAL D'UNE CUPULE. B1= PREMIERE BRACTEE, B2= SECONDE BRACTEE, B3= TROISIEME BRACTEE, F=FLEUR OU FRUIT EN DEVENIR, V=VALVE.

### A. Castaneoideae



### B. Fagoideae



= 3-flowered, 4-valved dichasium cupule; 
 = 2-flowered, 2-valved dichasium cupules with aborted lateral flowers; 
 = 3-flowered dichasium of flower cupules; 
 = 1-flowered, valveless acorn cupule.

FIGURE A5 : HYPOTHESES EVOLUTIVES DES CARACTERES MORPHOLOGIQUES ET REPRODUCTEURS (MANOS ET AL., 2001).

## A2.1 RECONSTRUCTIONS PHYLOGENETIQUES INTERGENRES

---

La structure et l'ontogenèse de la cupule ont suscité de très nombreuses interprétations phylogénétiques. La plus récente, sur laquelle se fonde la classification des Fagacées en genres, stipule que la cupule peut avoir deux origines différentes (Manos *et al.*, 2001).

Elle peut être constituée d'axes stériles à la base des inflorescences pistillées, appelées dichasium ; les valves de la cupule correspondent aux différentes ramifications de ces axes (Figure A4). De façon générale, le nombre de valves correspond au nombre de fleurs pistillées + 1 (Nixon and Crepet, 1989). La réduction du nombre de fleurs à une seule fleur centrale est apparue dans tous les genres. Une des hypothèses pour expliquer la formation du gland chez le genre *Quercus* serait que celui-ci dérive d'un dichasium ou les valves auraient fusionnées (Nixon and Crepet, 1989).

D'autres types apomorphes seraient dérivés du dichasium comme chez le genre *Chrysolepis* qui lui a des valves internes, ou le genre *Fagus* qui possède une cupule à 4 valves et 2 fleurs (Figure A5). L'alternative à la structure de type dichasium est la cupule florale, pour laquelle chaque fleur est sous tendue par une cupule sans valve, seul le genre *Lithocarpus* est concerné (Manos *et al.*, 2001; Oh and Manos, 2008).

Les analyses morphologiques des restes fossiles suggèrent que l'état ancestral (plésiomorphe) de la cupule correspond à la cupule de type dichasium et que l'état ancestral de la pollinisation correspond à la pollinisation entomophile.

En plaçant ces caractères sur un arbre phylogénétique (Figure A6) basé sur différents caractères morphologiques dont ceux évoqués précédemment, il est possible de tester les hypothèses évolutives de l'apparition de ces caractères en partant du principe que l'apparition d'un type de caractère est irréversible (Figure A7); ainsi, le syndrome pollinique anémophile serait apparu une fois dans l'évolution, tout comme la cupule florale, alors que l'apparition du type de fruit hypogé (les cotylédons restent dans le sol une fois la germination commencée) serait survenue deux fois au cours de l'évolution (Figure A7).

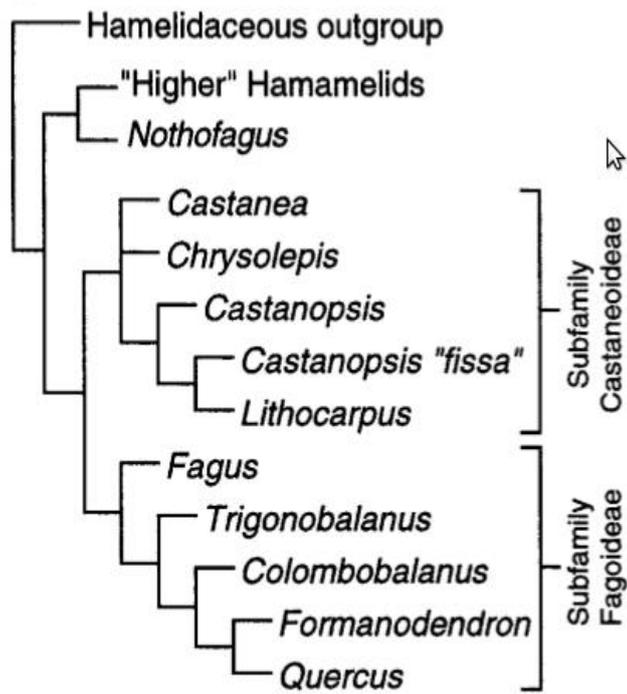


FIGURE A6 : REPRESENTATION PHYLOGENETIQUE HYPOTHETIQUE BASEE SUR DES CARACTERES MORPHOLOGIQUES (NIXON, 1985; NIXON AND CREPET, 1989; MANOS ET AL., 2001).

Les scénarios évolutifs concernant l'apparition des types de pollinisation et la formation des types de fruits dépendent entièrement de la phylogénie sur laquelle sont calqués ces caractères, l'utilisation de données moléculaires pour construire des arbres phylogénétiques s'avère donc primordiale pour confirmer ou réfuter les hypothèses évolutives émises.

Dans la suite de ce manuscrit, nous parlerons uniquement de l'évolution des phylogénies en fonction des marqueurs moléculaires utilisés, en omettant les implications sur les hypothèses concernant l'apparition des fruits, et ce, afin de ne pas alourdir le message.

## A2.2 RECONSTRUCTIONS PHYLOGENETIQUES A L'INTERIEUR DU GENRE QUERCUS

---

Très récemment, l'ornementation du grain de pollen a été proposé comme caractère morphologique pour reconstruire la phylogénie dans le genre *Quercus* (Denk and Grimm, 2009b). Les auteurs ont également été capables de donner une direction évolutive de l'ornementation du grain de pollen en définissant le groupe *Ilex* comme ancestral (Figure A8). En effet, ils ont remarqué que l'ornementation microrugulaire chez *Quercus ilex* était identique à celle du genre *Fagus* qui est plus ancien, ce caractère est reconnu comme plésiomorphe.

De plus, des études portant sur le développement des grains de pollen chez le genre *Fagus* et le genre *Quercus* ont montrées que le pattern trouvé chez *Quercus ilex* correspond au début du développement du sous-genre *Quercus s.l.* ayant une structure de pollen complexe comme chez le groupe des *Quercus s.s.* et le groupe des *Lobatae*.

Grâce à l'ornementation des grains de pollen par exemple, il a été possible de distinguer plusieurs grades correspondant aux *Cyclobalanopsis*, *Protobalanus*, *Lobatae* et *Quercus s.s.*. Le groupe *Ilex* est apparu comme étant très différent du groupe *Cerris*. L'exine des grains de pollen étant très résistante, l'utilisation de ce caractère sur des données fossiles facilitera grandement l'étude de la flore fossile qui est retrouvée en abondance.

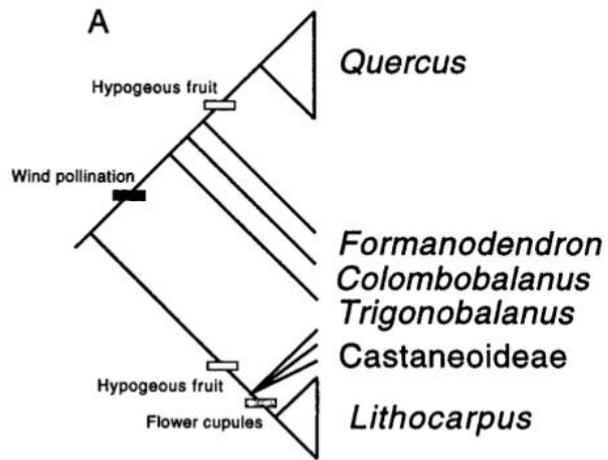


FIGURE A7 : ARBRE PHYLOGENETIQUE DE LA FAMILLE DES FAGACEES BASE SUR L'ANALYSE CLADISTIQUE DE LA MORPHOLOGIE ET SUR LEQUEL SONT DISTRIBUES LES CARACTERES DU FRUIT ET LE SYNDROME POLLINIQUE (FORMAN, 1966; NIXON, 1985; MANOS *ET AL.*, 2001).

### A3 PHYLOGENIES BASEES SUR DES CARACTERES MOLECULAIRES

---

Contrairement aux méthodes basées sur la parcimonie qui ne fait pas appel à un modèle d'évolution de séquences spécifique, l'utilisation de données moléculaires permet de faire appel aux méthodes phénétiques et probabilistes. La méthode phénétique (méthode basée sur des distances génétiques) quand elle est utilisée sur des caractères moléculaires, fait appel à différents modèles d'évolution de ces caractères. Cette méthode se base sur le calcul d'une distance globale entre paires d'individus. La construction d'un arbre phylogénétique est rapide, tient compte d'un modèle d'évolution mais c'est la totalité des caractères qui est prise en compte dans l'analyse en une seule fois (la distance entre individus).

La méthode probabiliste (faisant appel à des probabilités d'apparition de caractères) est utilisée sur des caractères moléculaires, elle fait appel à un modèle d'évolution et chaque site est pris individuellement. L'arbre inféré est assez long à obtenir, mais la plupart des analyses phylogénétiques modernes découle de ce type d'analyse.

TABLEAU A4 : CORRESPONDANCE ENTRE LES GROUPES TAXONOMIQUES DEFINIS PAR NIXON (NIXON, 1993), CAMUS (CAMUS, 1936-1954) ET MENITSKY (MENITSKY, 1984; MENITSKY, 2005) ET UTILISES DANS LA SUITE DE CETTE THESE.

Groupe défini dans la thèse	Classification Nixon	Classification Camus	Classification Menitsky
<i>Quercus s.s.</i>	Section <i>Quercus</i> sensu stricto	Section <i>Lepidobalanus</i> (sauf sous section <i>Ilex</i> ) + section <i>Mesobalanus</i> + section <i>Macrobalanus</i>	Sous-genre <i>Quercus</i>
<i>Lobatae</i>	Section <i>Lobatae</i>	Section <i>Erythrobalanus</i>	
<i>Protobalanus</i>	Section <i>Protobalanus</i>	Section <i>Protobalanus</i>	
<i>Cerris</i>			Sous-genre <i>Heterobalanus</i> Section <i>Heterobalanus</i>
<i>Ilex</i>			Sous-genre <i>Heterobalanus</i> Section <i>Ilex</i>

## A3.1 PHYLOGENIES BASEES SUR DES MARQUEURS CHLOROPLASTIQUES

---

### A3.1.1 CHEZ LES FAGACEES

---

A notre connaissance, il n'existe pas d'études publiées portant sur la phylogénie des Fagacées avec des marqueurs moléculaires chloroplastiques seuls autre que celle de Xu (Xu, 2004). Le marqueur utilisé est une zone intergénique entre le gène TrnD et le gène TrnT. Cette zone fait aux alentours de 1700 pb, et est coupée par 5 microsatellites (Xu, 2004).

La phylogénie obtenue par Xu à partir d'un fragment d'ADN chloroplastique à permis de retrouver les principaux genres de la famille des Fagacées (Figure A9) comme les *Fagus*, les *Lithocarpus*, les *Castanea*, les *Castanopsis* ainsi que les sous-groupes du genre *Quercus s.l.*; mais les relations phylogénétiques ne sont pas résolues à ce stade, puisque la forme de l'arbre obtenue correspond à une polytomie. L'utilisation d'autres types de marqueurs comme les marqueurs nucléaires s'avère plus que pertinente à ce stade.

### A3.1.2 CHEZ LE GENRE QUERCUS

---

Une des premières études basées sur des caractères moléculaires chloroplastiques chez le sous-genre *Quercus s.l.* est celle de Manos (Manos *et al.*, 1999) faite sur la base de sites de restrictions du génome chloroplastique entier (Figure A10).

Sur les 160 000 pb de l'ADN chloroplastique, 700 sites de restrictions ont été utilisés en parcimonie, ce qui correspond à 3,3 % du génome total. Cette analyse reconnaît cinq groupes mais sans relations évolutives entre eux: le groupe des chênes rouges (*Lobatae*), des chênes blancs (*Quercus s.s.*), des *Protobalanus*, du groupe *Cerris* et du groupe *Ilex*. L'arbre obtenu correspond à nouveau à une polytomie à cinq branches.

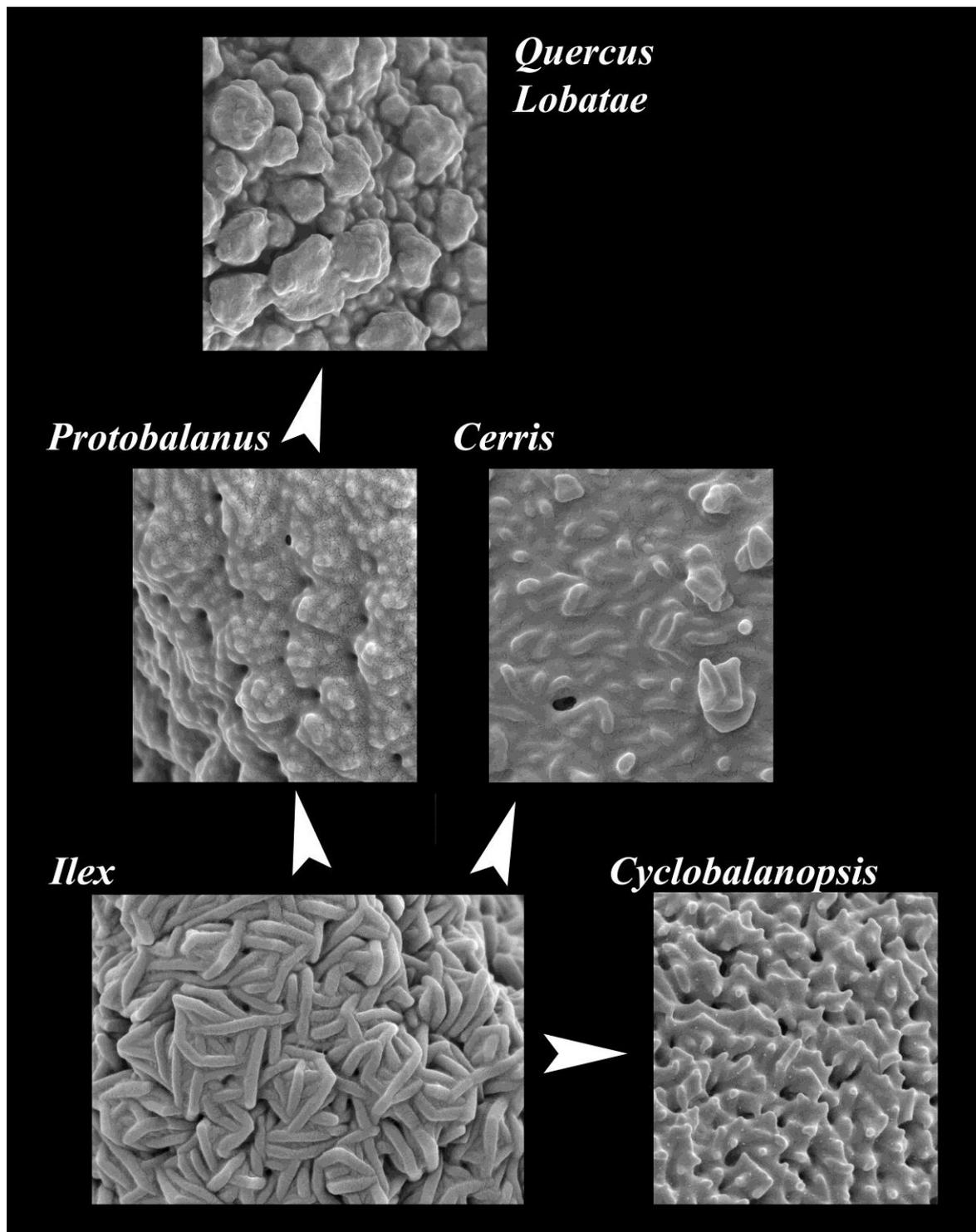


FIGURE A8 : HYPOTHESES EVOLUTIVES DES GROUPES DU GENRE QUERCUS SUITE A L'OBSERVATION DE L'ORNEMENTATION DES GRAINS DE POLLEN (DENK AND GRIMM, 2010).

Xu (Xu, 2004), en utilisant un fragment d'ADN chloroplastique de 1700 pb et en incluant l'information donnée par les marqueurs microsatellites à reconnu les mêmes groupes que l'étude précédente mais en éclatant le groupe des *Quercus s.s* en deux sous-groupes, en extrayant le groupe des Brachylepides (Camus, 1936-1954) du groupe *Cerris* et en établissant une certaine ancienneté des groupes *Cerris* et *Ilex* par rapport aux autres chênes (Figure A11).

Il est reconnu que l'ADN chloroplastique comporte bon nombre d'avantages du fait de son nombre de copies dans les cellules et de sa facilité d'extraction, mais il mute moins vite que l'ADN nucléaire. Cependant son hérédité uniparentale, exclusivement par la voie femelle, peut aboutir à l'appropriation du génome chloroplastique d'une espèce par une autre (phénomène de capture cytoplasmique) dans le cas d'hybridation.

Les études phylogénétiques portant sur des espèces interfertiles sont compromises par les processus agissant au niveau populationnel comme la capture cytoplasmique ou le tri incomplet des lignées ancestrales (Manos *et al.*, 1999). Il s'avère donc nécessaire d'utiliser d'autres types de marqueurs moléculaires comme les marqueurs nucléaires afin de pouvoir résoudre les relations phylogénétiques entre les différents groupes.

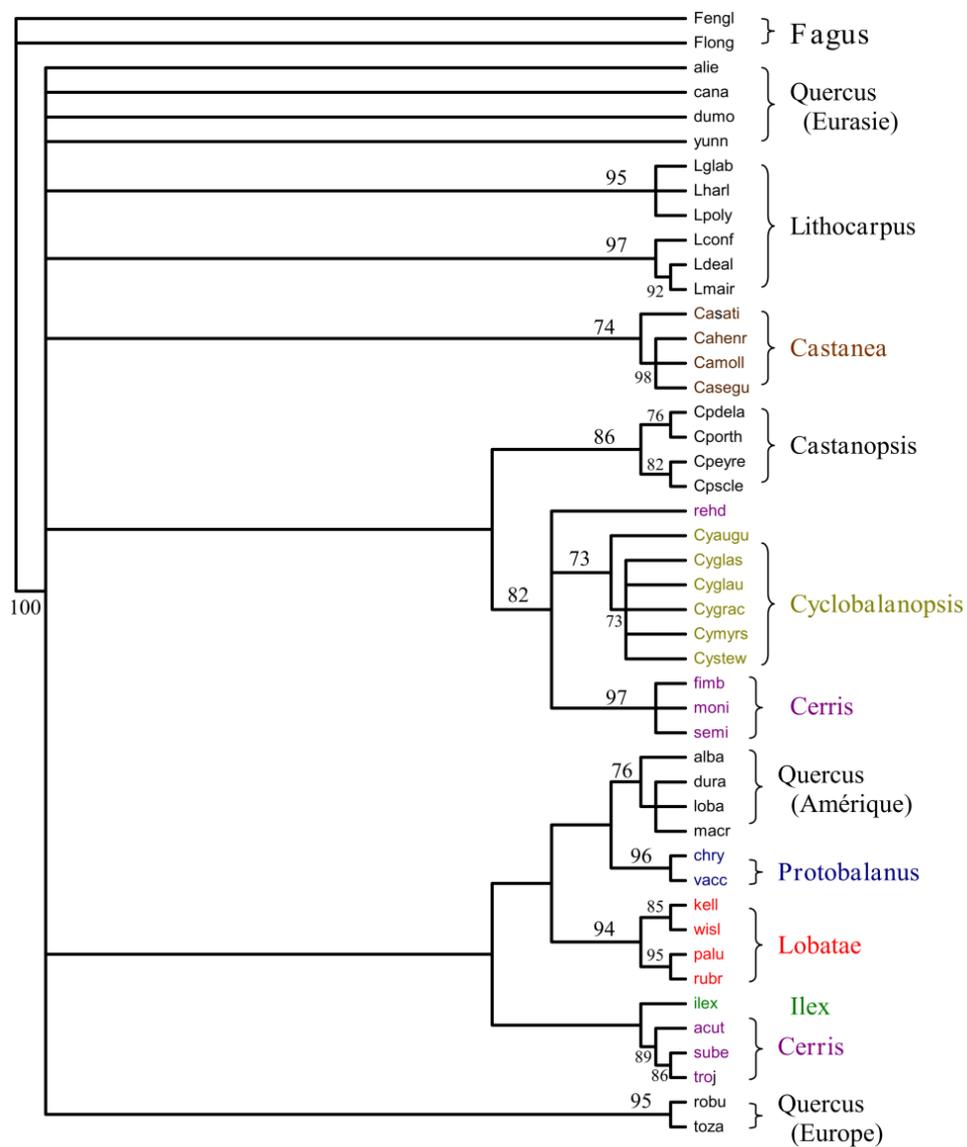


FIGURE A9 : PHYLOGENIE MOLECULAIRE BASEE SUR LE FRAGMENT CHLOROPLASTIQUE DT (XU, 2004). CET ARBRE EST L'ARBRE DE CONSENSUS STRICT BASE SUR TOUS LES ARBRES EQUI-PARCIMONIEUX. SEULS LES BOOTSTRAPS SUPÉRIEURS À 50% SONT INDICUÉS.

### A3.2 PHYLOGENIES BASEES SUR DES MARQUEURS NUCLEAIRES

---

Les marqueurs nucléaires utilisés pour les reconstructions phylogénétiques peuvent être classés en deux catégories : les marqueurs issus d'ADN Ribosomique comme les ITS (Internal Transcribed Spacers) ou les ETS (External Transcribed Spacers) qui se retrouvent en plusieurs copies dans le génome nucléaire ; et les marqueurs nucléaires qui se retrouvent en un faible nombre de copies dans le génome (low-copy nuclear genes). Une vingtaine de gènes ont été recensés très récemment dans la littérature (*Zimmer and Wen, 2013*).

Dans la suite de l'exposé, nous verrons que les marqueurs nucléaires utilisés dans les reconstructions phylogénétiques de la famille des Fagacées et du genre *Quercus*, sont essentiellement basés sur les ITS seuls avec l'apport occasionnel de données chloroplastiques (*Samuel et al., 1998; Manos et al., 1999; Manos and Stanford, 2001 ; Manos et al., 2001; Bellarosa et al., 2005; Oh and Manos, 2008; Denk and Grimm, 2009 , 2010*), et anecdotiquement, sur un gène nucléaire présent en une seule copie qui vient en complément des ITS (*Oh and Manos, 2008*). Enfin, une méthode prospectant l'ensemble du génome nucléaire basée sur des AFLP (Amplified Fragment Length Polymorphism) a été récemment proposée (*Pearse and Hipp, 2009*).

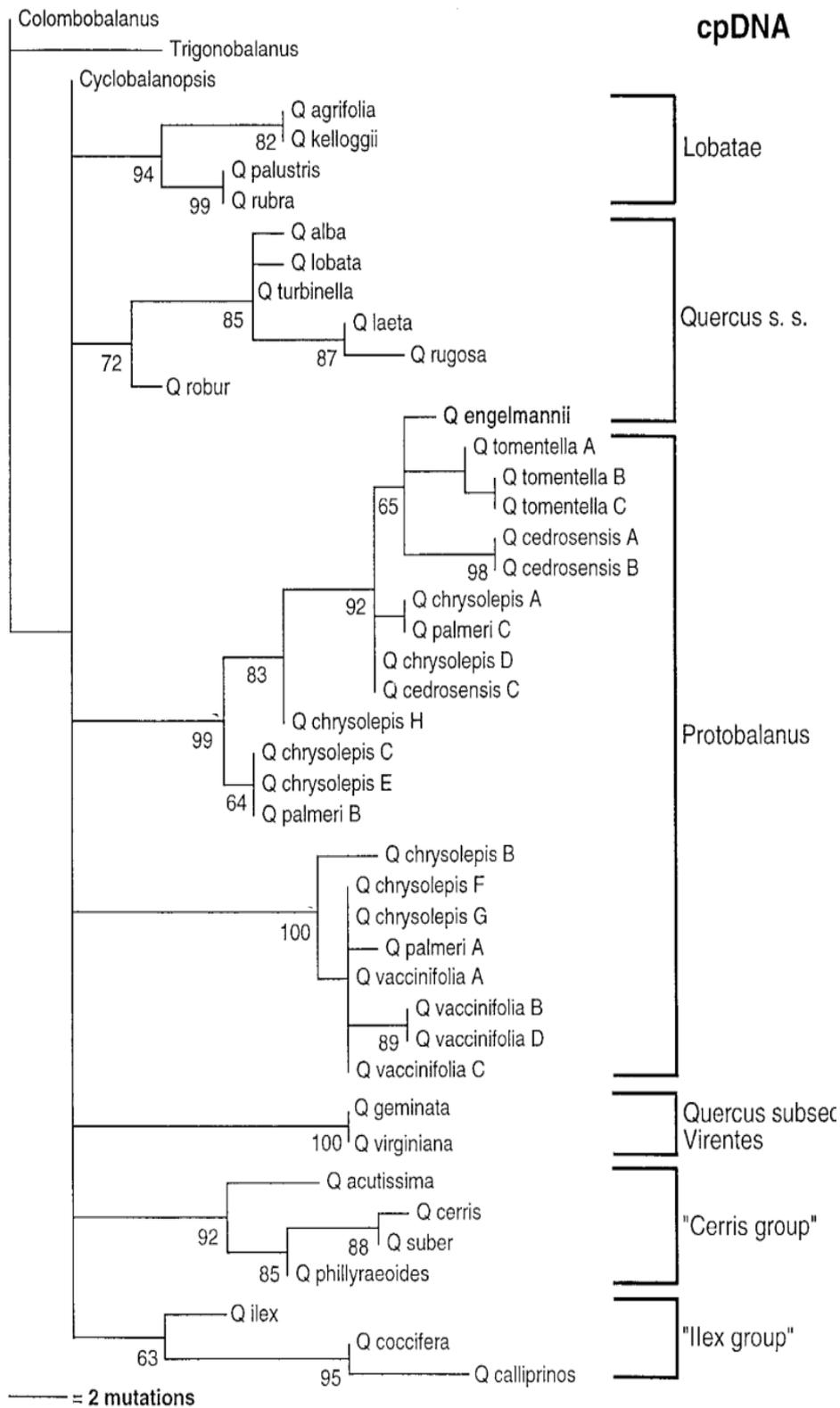


FIGURE A10 : PHYLOGENIE MOLECULAIRE DU SOUS-GENRE *QUERCUS*, BASEE SUR LES SITES DE RESTRICTION DE L'ADN CHLOROPLASTIQUE (MANOS ET AL., 1999). CET ARBRE EST L'ARBRE LE PLUS PARCIMONIEUX, LES SUPPORTS DONNES SONT LES BOOTSTRAPS SUPERIEURS A 50 %.

---

### A3.2.1 CHEZ LES FAGACEES

---

Trois études ont porté sur les relations phylogénétiques dans la famille entière (Manos *et al.*, 2001; Oh and Manos, 2008; Denk and Grimm, 2010). Deux d'entre elles sont basées uniquement sur les ITS (Manos *et al.*, 2001; Denk and Grimm, 2010) et la troisième combine les ITS et un marqueur nucléaire représenté en une seule copie : le CRC, ou CRABS CLAW (Oh and Manos, 2008) qui est un gène impliqué dans le développement des carpelles (c'est une enveloppe de protection des ovules). Ce gène a été choisi par l'équipe de Manos pour son rôle fonctionnel dans le développement d'un organe sur lequel portait des inférences phylogénétiques (voir paragraphe précédent). Ces trois études arrivent toutes à reconnaître les différents genres sans ambiguïté (Figures A12, A13 et A14) mais les relations entre elles sont les plus nettes en combinant les données de l'ITS à celles du CRC (Figure A14). Ainsi, en racinant sur le genre *Fagus*, il apparaît que le groupe le plus proche de *Fagus* soit composé des genres *Colombobalanus*, *Formanondendron* et *Trigonobalanus* dans chacune des 3 analyses (Figures A12, A13 et A14).

Les reconstructions basées sur l'ITS seul et utilisant la parcimonie (Figure A12) ne sont pas capables de bien séparer les genres à l'exception peut être de la formation d'un groupe composé par le genre *Castanea* et le genre *Castanopsis*.

Celles basées sur l'ITS mais utilisant le maximum de vraisemblance (Figure A13) suggèrent une certaine ancienneté des *Castanopsis*, puis des *Castanea*, des *Lithocarpus*, des *Chrysolepis*, des *Notholithocarpus densiflorus* (qui s'appelaient *Lithocarpus densiflorus* jusqu'en 2008) et du genre *Quercus*, mais aucune relation entre les genres n'apparaît à ce stade.

L'utilisation conjointes des séquences des ITS et de CRC (Figure A14) est capable de regrouper entre eux les genres *Chrysolepis* et *Lithocarpus* à la suite du groupe composé par *Colombobalanus*, *Formanondendron* et *Trigonobalanus*; puis les genres *Castanea* et *Castanopsis* (comme suggéré par l'analyse en parcimonie sur les ITS), puis les *Notholithocarpus* et enfin le genre *Quercus*.

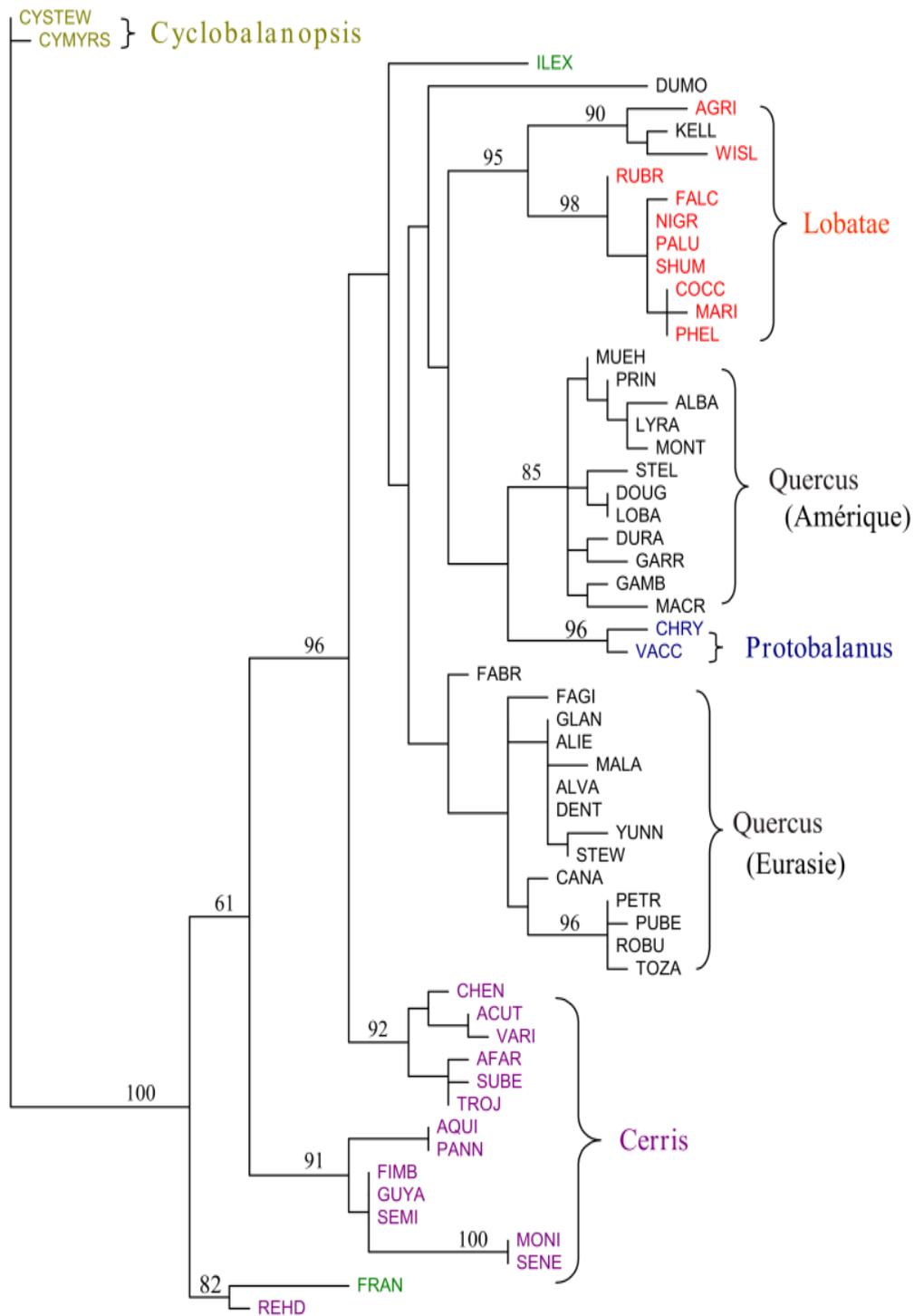


FIGURE A11 : UN DES 1000 ARBRES LES PLUS PARCIMONIEUX BASE SUR LE FRAGMENT DT (XU, 2004). SEULS LES BOOTSTRAPS SUPERIEURS A 50% SONT INDIQUES.

Les deux analyses utilisant le maximum de vraisemblance (Figures A13 et A14) indiquent que le genre *Quercus* apparaît comme étant le dernier de la famille à avoir divergé, juste après les *Notholithocarpus*.

---

### A3.2.2 CHEZ LE GENRE *QUERCUS*

---

De nombreuses études ont tenté avec plus ou moins de succès de comprendre l'organisation des groupes du genre *Quercus* (Samuel *et al.*, 1998; Manos *et al.*, 1999; Manos and Stanford, 2001a; Manos *et al.*, 2001; Oh and Manos, 2008; Pearse and Hipp, 2009; Denk and Grimm, 2010).

Un sérieux écueil quand à l'utilisation des ITS en phylogénie a été révélé en comparant l'étude de Samuel (Samuel *et al.*, 1998) à celle de Manos (Manos *et al.*, 1999).

En effet, ces deux études sont arrivées à des divergences très importantes : l'étude de Samuel (Figure A15) suggérant que le genre *Quercus* serait scindé par un groupe composé d'espèces appartenant aux groupes des *Cerris* (*Quercus suber* + *Q. macrolepis* + *Q. acutissima*) et au groupe *Lobatae* (*Quercus rubra*), et un groupe composé des espèces des clades *Ilex* (*Quercus Ilex* + *Q. coccifera*), *Cerris* (*Quercus cerris*) et *Quercus s.s* (*Quercus virginiana* + *Q. petraea* + *Q. robur*).

Manos (Manos *et al.*, 1999), quant à lui, regroupe effectivement les groupes *Cerris* et *Ilex* en un seule clade (Figure A16), mais sépare bien les *Lobatae* des *Quercus s.s* (Figure A16). Ces deux études sont réalisées toutes les deux sous l'hypothèse de parcimonie et la seule différence, sans parler de l'échantillonnage qui n'est pas identique provient donc des ITS qui ont été amplifiés et séquencés. Ce problème a fait l'objet d'une mise au point (Mayol and Rossello, 2001) ou les auteurs sont arrivés à la conclusion que l'équipe de Samuel a séquencé des ITS d'origine paralogue, c'est-à-dire qu'il y a eu duplication de ces ITS dans le génome puis perte de certains d'entre eux. Par la suite nous ne ferons donc plus référence aux travaux de Samuel *et al.* (1998).

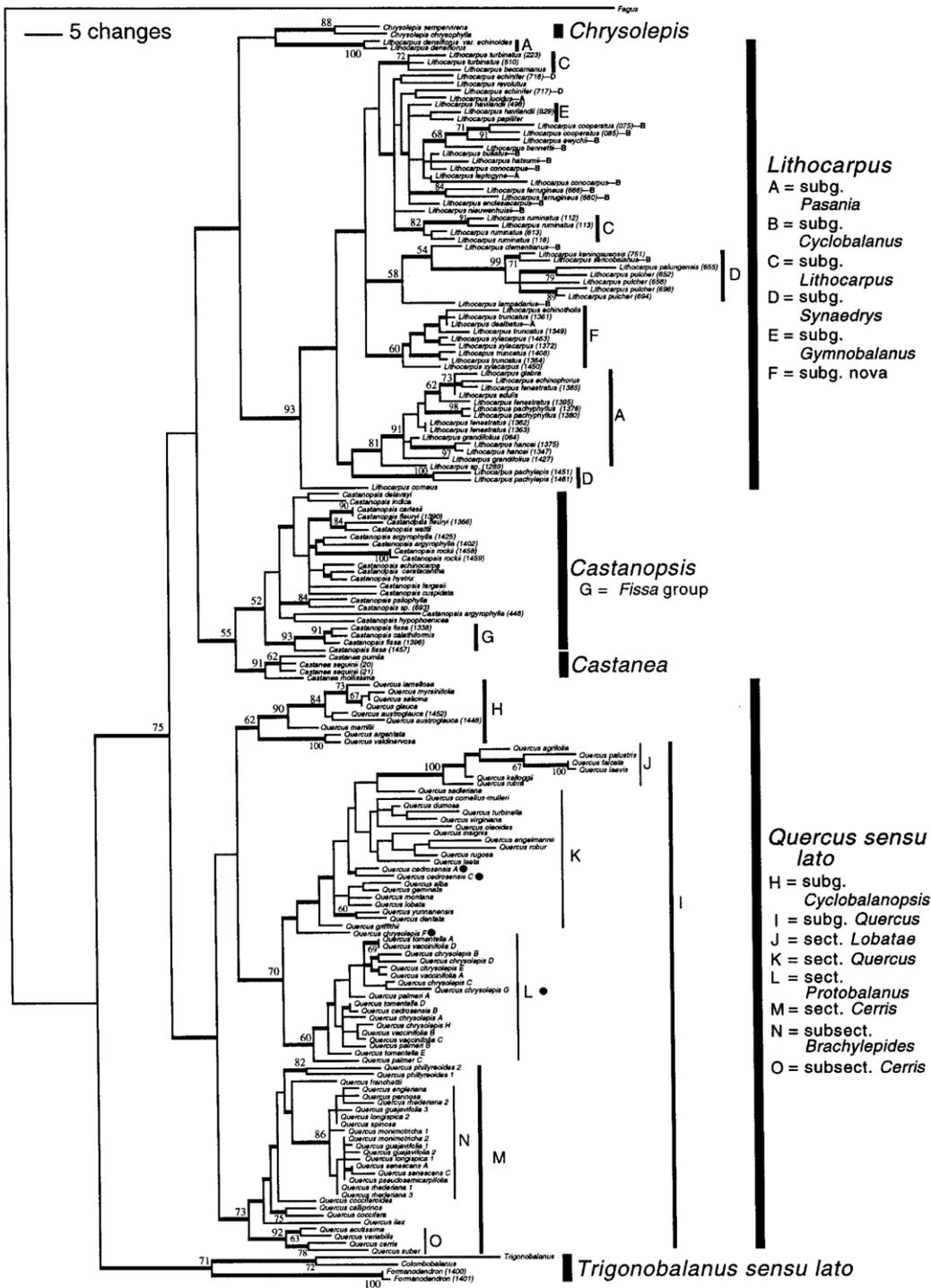


FIGURE A12 : UN DES ARBRES LES PLUS PARCIMONIEUX BASE SUR UN JEU DE DONNEES COMPOSE DE 179 TAXONS SUR LE MARQUEUR NUCLEAIRE ITS (MANOS ET AL., 2001). SEULS LES BOOTSTRAPS SUPERIEURS A 50% SONT INDQUES.

Afin de simplifier la lecture au maximum et en espérant que le message principal passe facilement, je vais décrire dans la suite du texte les éléments principaux des études qui seront présentées dans l'ordre de publication.

Les *Cyclobalanopsis*, les *Cerris* et les *Ilex* seront décrits de façon indépendante des *Protobalanus*, des *Lobatae* et des *Quercus s.s.* qui se retrouvent ensemble quelles que soient les études.

---

#### ETUDE DE 1999 (MANOS ET AL., 1999) :

---

Le but de cette recherche est d'étudier la phylogénie, la biogéographie ainsi que les processus impliqués dans la différenciation moléculaire du genre *Quercus*.

La reconstruction phylogénétique proposée repose sur l'utilisation unique des ITS (Figure A16 et Tableau A5) ou sur la combinaison de ces données avec les sites de restriction de l'ADN chloroplastique (Tableau A5).

L'arbre est enraciné sur un individu du genre *Colombobalanus* et du genre *Trigonobalanus* (Figure A16).

Le résultat à retenir est que le groupe des *Cyclobalanopsis* aurait divergé avant le reste du genre *Quercus* (Figure A16), ceci étant en accord avec la classification de Camus (Camus, 1936-1954) qui scindait le genre *Quercus* en un sous-genre *Cyclobalanopsis* et un sous-genre *Euquercus* qui regroupe les autres espèces.

Le groupe qui aurait divergé par la suite est constitué du groupe des *Cerris* et du groupe des *Ilex* (Figure A16) ; il est intéressant de noter que dans cette étude, les auteurs reconnaissent un seul groupe –le groupe des *Cerris*- en regroupant toutes ces espèces entre elles, alors que le groupe des *Ilex* peut déjà être distingué.

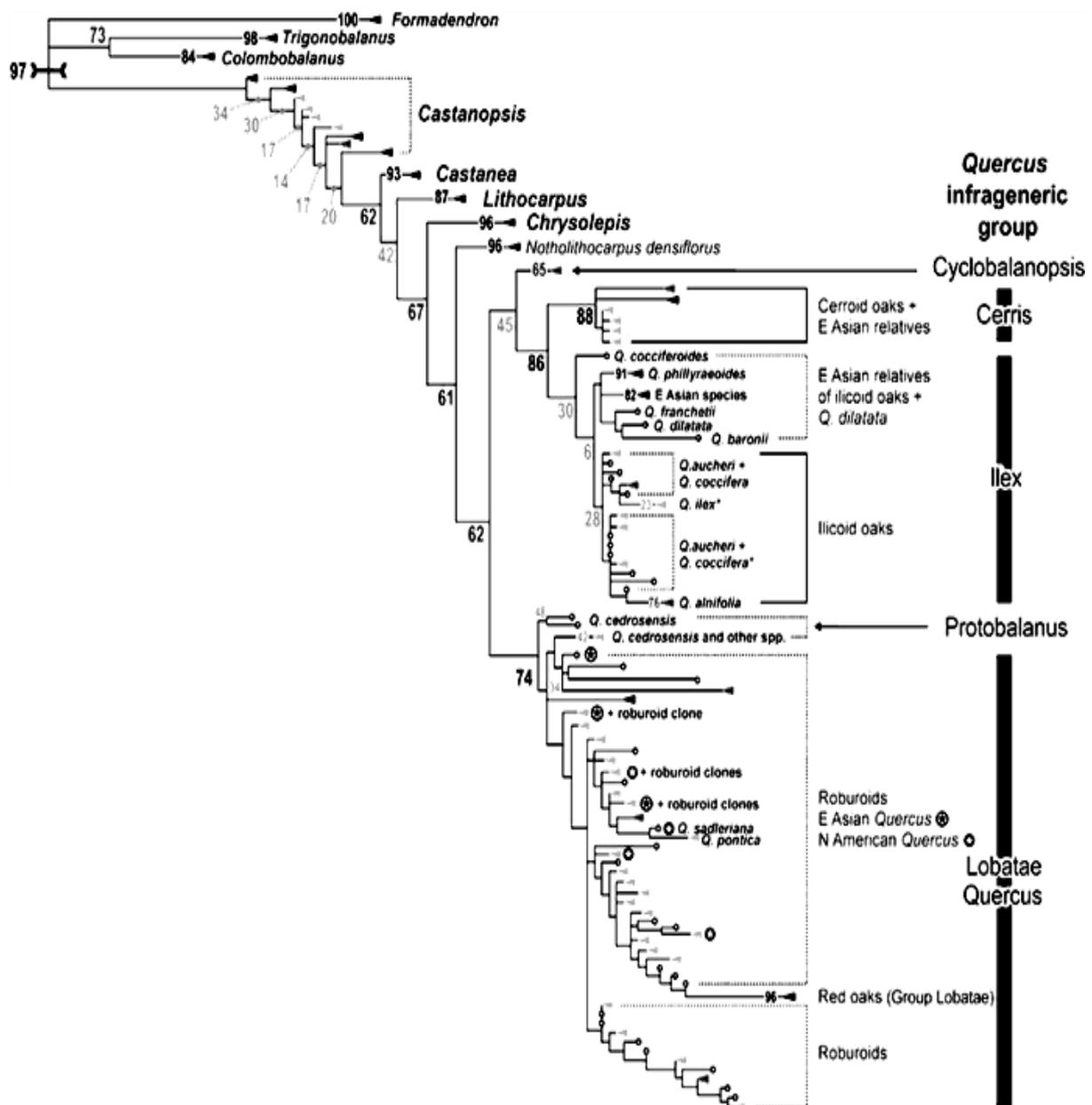


FIGURE A13 : ARBRE OBTENU EN MAXIMUM DE VRAISEMBLANCE SUR TOUS LES ITS PUBLIES ET CEUX AJOUTES LORS DE CETTE ETUDE (DENK AND GRIMM, 2010).TOUS LES SUPPORTS SONT INDIQUEES BIEN QUE CEUX SUPERIEURS A 50 % SOIENT DE COULEUR PLUS FONCEE.

Ensuite, le groupe des *Lobatae* composé de quatre individus (Tableau A5), puis le groupe des *Protobalanus* et des *Quercus s.s.* (Figure A16) dérivent à la suite des groupes précédents. Les auteurs scindent le genre *Quercus* en deux clades qu'ils dénomment « Old World Oaks » pour les groupes *Cyclobalanopsis* + *Cerris* + *Ilex* et les « New World Oaks » pour les *Protobalanus* + *Quercus s.s.* + *Lobatae*. Les espèces du clade « Old World Oaks » sont assemblées en un sous-groupe composé des *Cerris* + *Ilex* et un sous-groupe de *Cyclobalanopsis* ; les espèces du groupe « New World Oaks » sont scindées en un sous-groupe *Protobalanus*+*Quercus s.s.* ; et un sous-groupe avec uniquement des *Lobatae* (Figure A16). Il faut cependant noter que l'emploi du terme « New World » pour les *Quercus s.s.* eurasiatiques peut paraître inapproprié puisque ceux-ci se retrouvent également en Eurasie.

---

#### ETUDE DE 2001 (MANOS ET AL., 2001) :

Cette étude a comme but de fournir une phylogénie des principaux genres composant la famille de Fagacées dans le but d'étudier l'évolution des principaux caractères morphologiques dont nous avons parlé précédemment ; ceci explique pourquoi le genre *Quercus* est aussi bien représenté en nombre d'espèces que les autres genres (Tableau A5 et Figure A12).

La méthode de reconstruction utilisée est comme précédemment la parcimonie avec l'utilisation d'ITS comme marqueurs moléculaires mais un échantillonnage différent (Tableau A5 et Figure A12).

Cette fois-ci, il y a inversion dans l'ordre des *Cyclobalanopsis* qui auraient divergés après le groupe *Cerris*+*Ilex* (Figure A12), puis les *Protobalanus* et les *Lobatae* + *Quercus s.s.* alors que dans l'étude précédente, les *Protobalanus* se plaçaient avec les *Lobatae* (Figure A16). Remarquons également que les *Cerris* et *Ilex* n'apparaissent pas individuellement comme deux groupes monophylétiques, alors qu'ils l'étaient dans l'étude précédente.

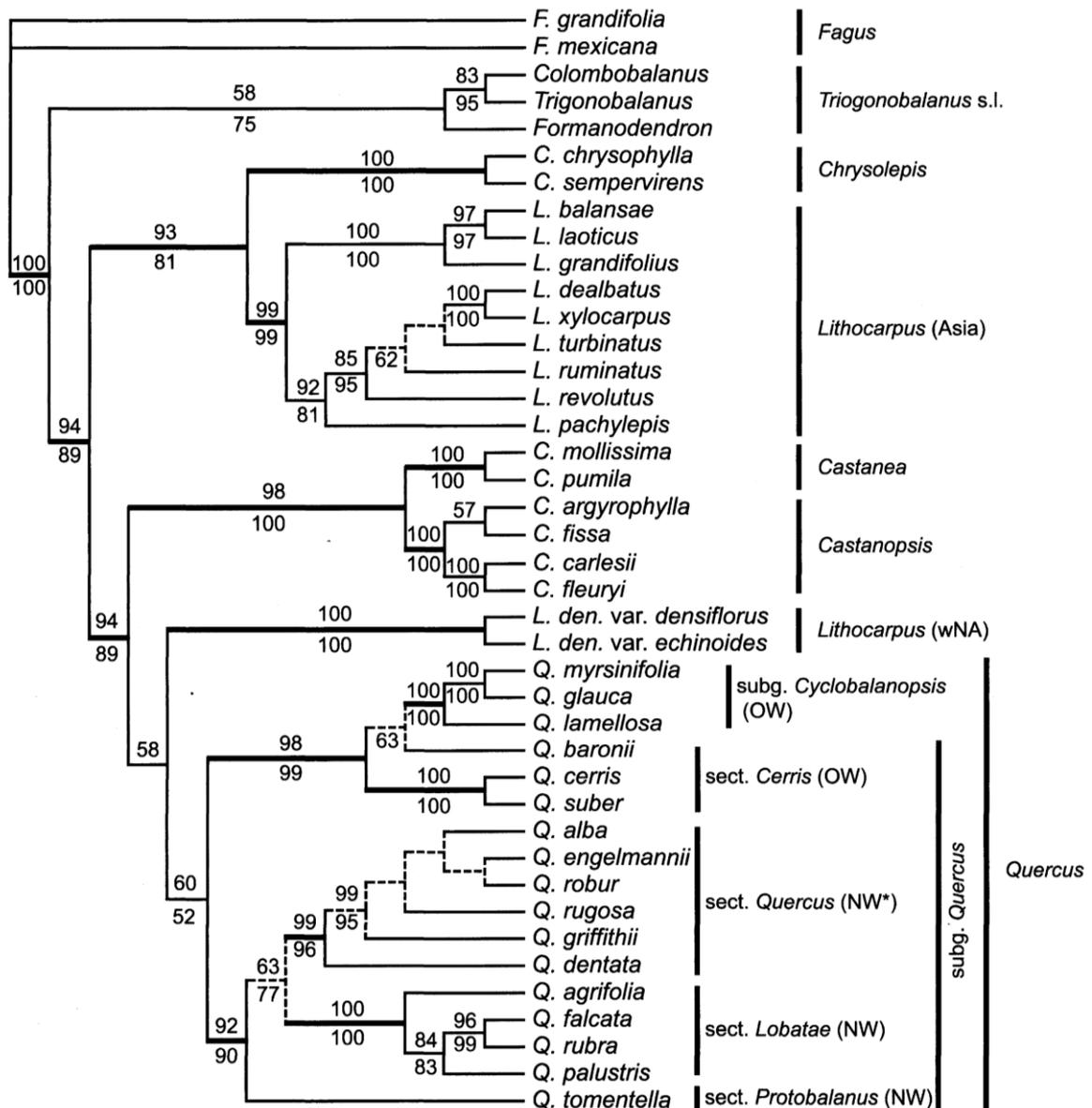


FIGURE A14 : ARBRE OBTENU EN MAXIMUM DE VRAISEMBLANCE COMBINANT LES DONNEES ISSUES DES ITS ET DU GENE CRC (OH AND MANOS, 2008).

LES BOOTSTRAPS INDIQUES AU DESSUS DES BRANCHES PROVIENNENT DE L'ANALYSE EN MAXIMUM DE PARCIMONIE, LES BOOTSTRAPS SOUS LA BRANCHE DE LA RECHERCHE DE L'ARBRE LE PLUS VRAISEMBLABLE.

#### ETUDE DE 2001 (MANOS AND STANFORD, 2001 ) :

---

Une autre étude basée comme précédemment sur les ITS a été menée dans le but d'étudier la biogéographie historique des différents genres de la famille. Pour ce faire, les ITS ont été utilisés avec comme racine, une espèce de *Chrysolepis*, de *Castanopsis*, de *Castanea* et une de *Lithocarpus* (Tableau A5 et Figure A17). Cette fois ci, une nouvelle topologie apparaît avec la formation de deux groupes monophylétiques composés par les différents membres des espèces du clade « Old World Oaks » et des espèces appartenant au clade « New World Oaks » (Figure A17).

Au sein du groupe « Old World Oaks », les *Cyclobalanopsis* sont séparés des *Cerris* + *Ilex* (Figure A17). Comme précédemment ces deux derniers groupes ne sont pas séparés en groupes monophylétiques

#### ETUDE DE 2005 (BELLAROSA ET AL., 2005) :

---

L'utilisation de l'ITS comme marqueur moléculaire a été confirmée lors d'une étude à une échelle locale puisqu'elle est basée uniquement sur 11 espèces de *Quercus* présentes en Italie (Tableau A5 et Figure A18). Il y a donc uniquement dans ce jeu de données des espèces appartenant aux groupes *Cerris*, *Ilex* et *Quercus s.s.* (les autres groupes n'étant pas présents en Italie).

En racinant sur une espèce du genre *Colombobalanus* et une espèce du genre *Trigonobalanus*, le groupe *Ilex* se retrouve en groupe frère des *Cerris* (Figure A18), les *Quercus s.s.* étant un groupe frère de ces derniers. Les deux groupes *Ilex* et *Cerris* forment deux groupes monophylétiques.

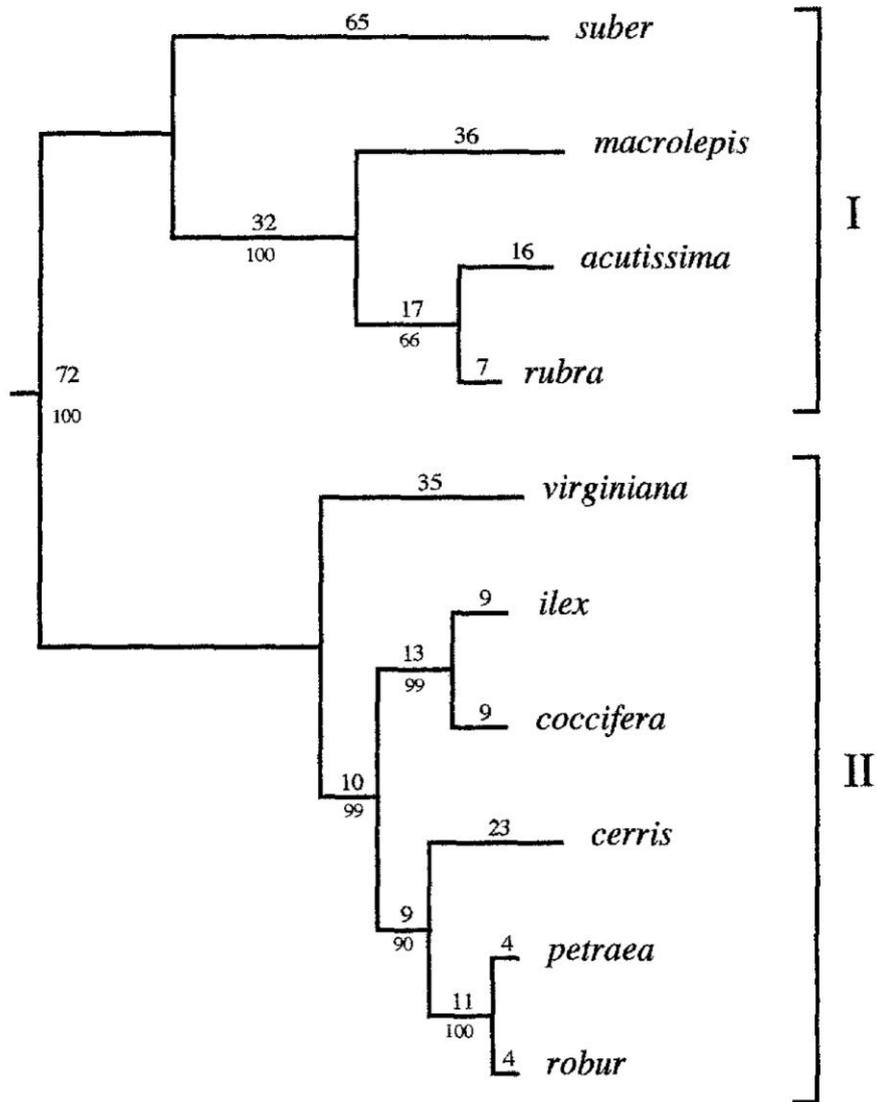


FIGURE A15 : ARBRE OBTENU EN PARCIMONIE ET BASE SUR LES ITS DE 10 ESPECES DU GENRE *QUERCUS* (SAMUEL *ET AL.*, 1998). TOUTES LES VALEURS DE BOOTSTRAPS SONT REPRESENTEES SOUS LES BRANCHES ALORS QUE LES VALEURS AU DESSUS SONT LE NOMBRE DE BASES QU'IL FAUT MODIFIER POUR PASSER D'UNE ESPECE A UNE AUTRE.

#### ETUDE DE 2008 (MANOS ET AL., 2008) :

---

Cette étude est sans doute une des plus riches car elle inclue en plus des ITS, un gène nucléaire supplémentaire (CRC), portant à plus de 3000 pb l'alignement total (Tableau A5).

L'utilisation du maximum de vraisemblance au jeu de données ainsi constitué permet de retrouver les deux groupes « Old World Oaks » et « New World Oaks » avec de fortes valeurs de bootstraps aux branches (Figure A14).

Deux petites remarques sont importantes à relever au niveau de l'échantillonnage. En effet, dans cette étude, une seule espèce du groupe *Ilex* est échantillonnée (*Quercus baronii*, Figure A14) et une seule espèce de *Protobalanus* aussi (*Quercus tomentella*, Figure A14). Le genre *Notholithocarpus* apparaît comme étant le groupe frère le plus proche du genre *Quercus* (Figure A14), mais le genre *Quercus* apparaît comme peu soutenu avec un bootstrap de 52% (Figure A14) ; cette remarque est aussi applicable au groupe *Notholithocarpus* + *Quercus* qui lui est inférieur à 50% (Figure 14) ; ceci provenant en partie du signal phylogénétique faible des ITS de cette étude (pour faciliter la lecture, ces données ne sont pas incluses dans le présent manuscrit).

#### ETUDE DE 2009 (PEARSE AND HIPPI, 2009) :

---

Dans le but de tester l'hypothèse selon laquelle des plants de chênes non présents naturellement d'un endroit seraient moins attaqués que l'espèce locale (les prédateurs étant adaptés aux espèces locales), les auteurs ont établi une phylogénie moléculaire basée sur les AFLP (amplified fragment-length polymorphism). L'ADN génomique est digéré par une ou plusieurs enzymes de restrictions, des adaptateurs sont ensuite placés et ligés aux sites de coupures puis une réaction d'amplification (PCR) est réalisée. A la fin, un individu se verra attribué plusieurs bandes de tailles différentes, et la présence ou l'absence de ces bandes par rapport aux autres individus constituera le polymorphisme étudié.

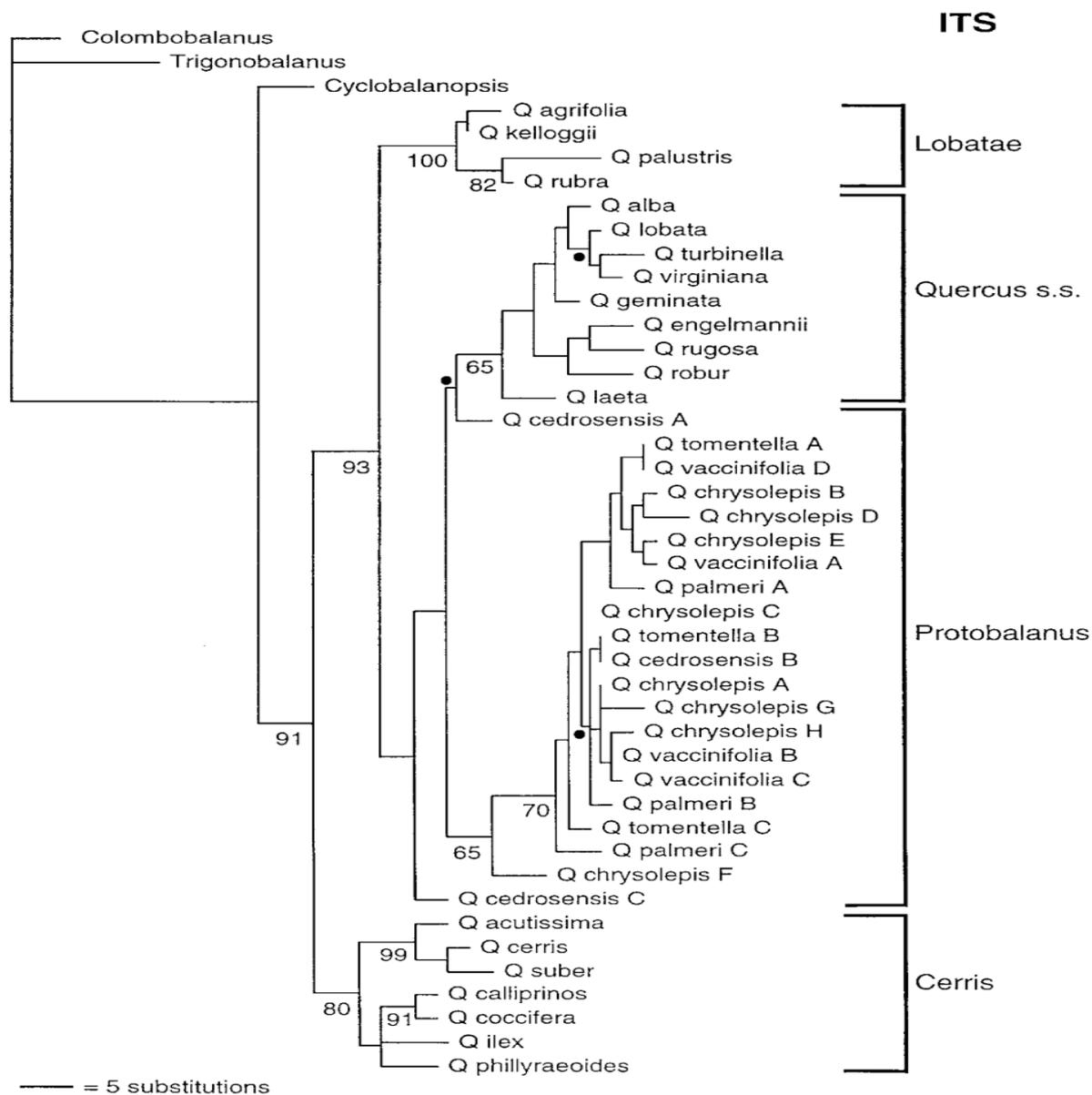


FIGURE A16 : PHYLOGRAMME OBTENU EN PARCIMONIE ET BASE SUR LES ITS DE DIFFERENTES ESPECES DU GENRE *QUERCUS* (MANOS ET AL., 1999). TOUTES LES VALEURS DE BOOTSTRAPS SUPERIEURES A 50% SONT REPRESENTEES.



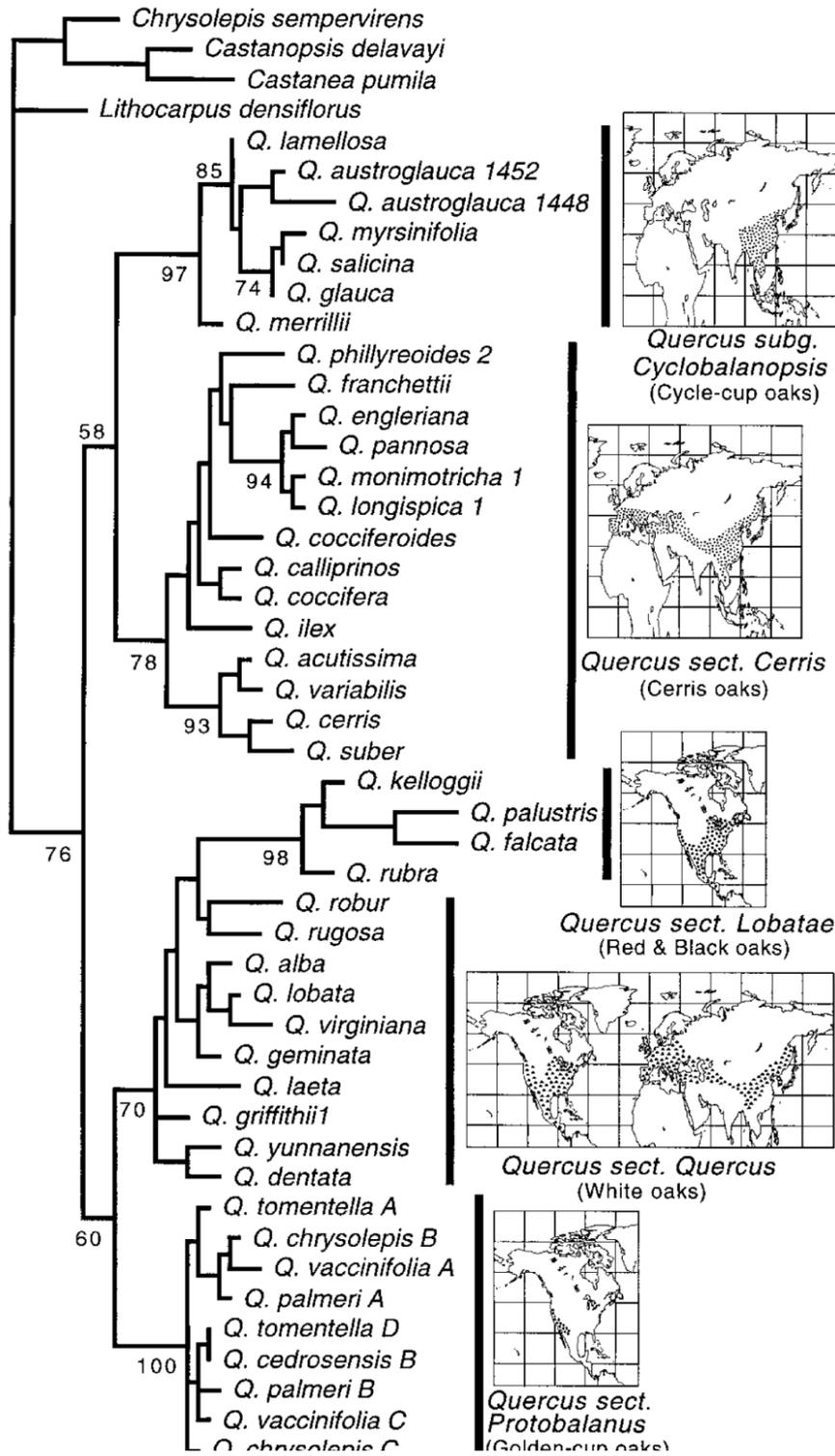


FIGURE A17 : PHYLOGRAMME OBTENU EN PARCIMONIE ET BASE SUR LES ITS DE DIFFERENTES ESPECES DU GENRE *QUERCUS* (MANOS AND STANFORD, 2001A). TOUTES LES VALEURS DE BOOTSTRAPS SUPERIEURES A 50% SONT INDIQUEES.

Une analyse Bayésienne est ensuite appliquée sur les distances entre les individus à partir de la matrice présence/absence d'amplification (Figure A19), cet arbre est raciné sur *Castanea*.

D'après cet arbre, le sous-groupe « Old World Oaks» est composé de deux espèces du groupe des *Cyclobalanopsis* qui sont frères des *Cerris* et des *Ilex* (Figure A19), mais ce groupe est assez faiblement soutenu (probabilité postérieure de 0.64).

Le sous-groupe « New World Oaks » est composé par le groupe des *Lobatae* qui se place en groupe frère des *Protobalanus* + *Quercus s.s*, et ce, de façon assez peu soutenue également (probabilité postérieure de 0.78).

---

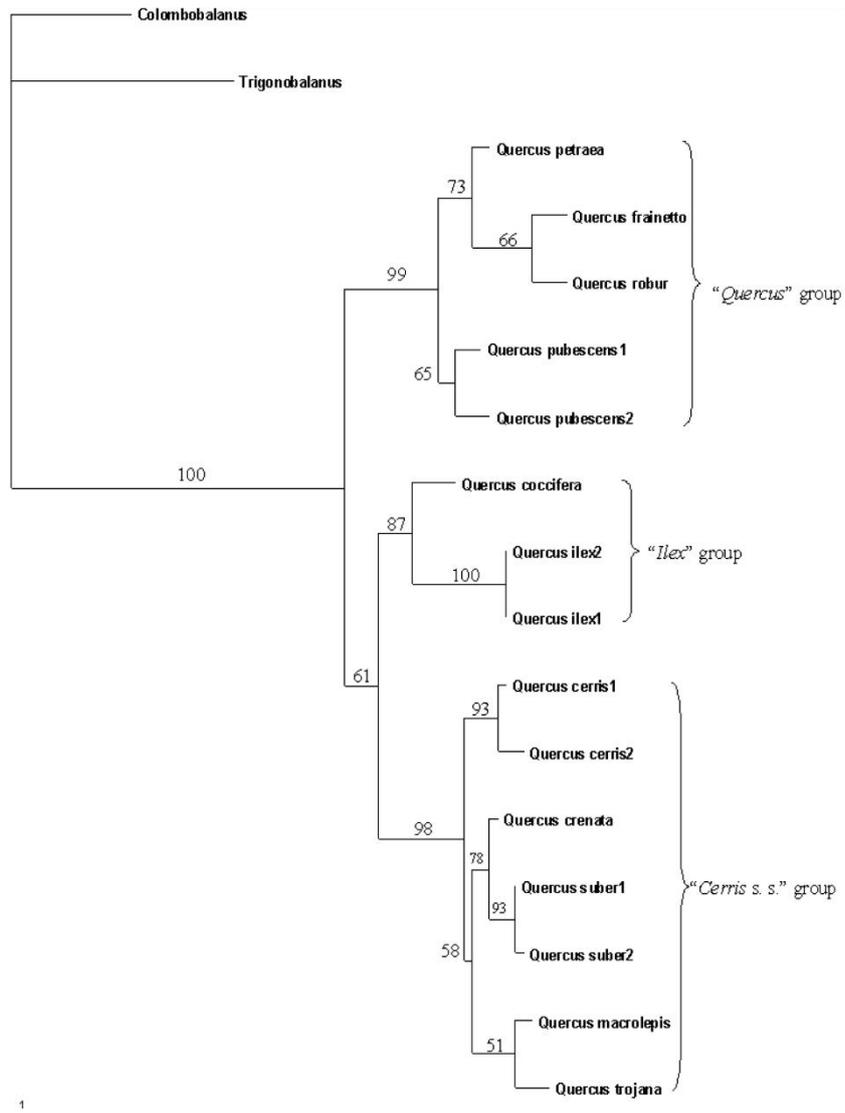
#### ETUDE DE 2010 (DENK AND GRIMM, 2010) :

---

Cette étude est uniquement basée sur des ITS mais sur tous les ITS présents dans les banques de données plus celles ajoutées par les auteurs (après avoir soustrait les séquences supposées paralogues pour éviter le problème rencontré dans l'étude dont nous avons parlé au début de ce chapitre).

Il est intéressant de noter que tous les genres (à l'exception de *Castanopsis* et du genre *Quercus*) ont des supports de branches élevés (Figure A13 et Tableau A5) mais que la résolution (c'est-à-dire l'enchaînement des genres entre eux) est faible car les supports de branches ne dépassent jamais les 70 % mais stagne autour des 60 %.

Le genre *Notholithocarpus* se trouve de nouveau être le groupe le plus proche du genre *Quercus* avec un support de 61 % (Figure A13). Le genre *Quercus* est supporté par une faible valeur de 62 % (Figure A13), lui-même étant scindé en deux sous-groupes : le sous-groupe « New World Oaks» et le sous-groupe « Old World Oaks».



1

FIGURE A18 : UN DES 4 ARBRES LES PLUS PARCIMONIEUX ISSUS DE L'ANALYSE ITS D'ESPECES ITALIENNES (BELLAROSA *ET AL.*, 2005). TOUTES LES VALEURS DE BOOTSTRAPS SONT REPORTEES SUR LA FIGURE.

Le sous-groupe « Old World Oaks » est très peu soutenu avec un support de branche de 45%, il contient le groupe des *Cyclobalanopsis* qui est frère du groupe *Cerris* + *Ilex* (les espèces du groupe *Ilex* ayant été rajoutées par les auteurs) ; le groupe *Ilex* étant peu soutenu (Figure A13). Le sous-groupe « New World Oaks » est bien soutenu avec une valeur de branche de 74% mais les *Protobalanus*, les *Quercus s.s* ; et les *Lobatae* sont mal individualisés (Figure A13).

Pour conclure, nous allons revenir sur les concordances générales et les discordances que nous avons trouvées dans les différentes études.

Les *Protobalanus*, les *Lobatae* et les *Quercus s.s* se retrouvent toujours ensemble quels que soient la méthodologie employée et les marqueurs utilisés (Manos *et al.*, 1999; Manos and Stanford, 2001 ; Manos *et al.*, 2001; Manos *et al.*, 2008; Pearse and Hipp, 2009; Denk and Grimm, 2010).

Cependant, les *Protobalanus* peuvent se retrouver en groupe frère des *Lobatae+Quercus s.s* (ceux-ci étant bien différenciés comme avec le CRC+ITS ou mélangés avec les ITS), ou bien les *Lobatae* se retrouvent en groupe frère des *Protobalanus + Quercus s.s* (Manos *et al.*, 1999; Pearse and Hipp, 2009).

Chez les groupes *Ilex*, *Cerris* et *Cyclobalanopsis*, les relations sont encore plus compliquées. Les groupes *Ilex* et *Cerris* peuvent se retrouver mélangés (Manos and Stanford, 2001a; Oh and Manos, 2008) ou bien différenciés et/ou associés fortement entre eux (Manos *et al.*, 1999; Manos *et al.*, 2001; Bellarosa *et al.*, 2005; Pearse and Hipp, 2009; Denk and Grimm, 2010).

Le groupe des *Cyclobalanopsis* est toujours bien dissocié mais sa place varie entre les études : ils auraient divergés après les *Ilex+Cerris* (Manos *et al.*, 2001), ou avant les *Ilex+Cerris* (Manos *et al.*, 1999) ou en groupe frère des *Ilex+Cerris* (Manos and Stanford, 2001b; Oh and Manos, 2008; Pearse and Hipp, 2009; Denk and Grimm, 2010).

Nous pourrions résumer en déclarant qu'il existe un consensus entre les études pour garder séparés les deux sous-groupes « Old World Oaks » et « New World Oaks » mais la hiérarchie de ces sous-groupes et leur composition n'est pas stable.

TABLEAU A5 : TABLEAU RECAPITULANT LES RESULTATS OBTENUS LORS DE DIFFERENTES ETUDES SUR LE GENRE *QUERCUS* (MANOS *ET AL.*, 1999; MANOS AND STANFORD, 2001A; MANOS *ET AL.*, 2001; BELLAROSA *ET AL.*, 2005; OH AND MANOS, 2008; DENK AND GRIMM, 2010). LES NUMEROS DES FIGURES CORRESPONDANTES SONT INSCRITES SOUS LE NOM DE L'EQUIPE QUI A MENE L'ETUDE, LE NOM DES MARQUEURS, LEUR TAILLE ET LA METHODE DE RECONSTRUCTION PHYLOGENETIQUE EST INDIQUEE. SOUS CHAQUE NOM DE GENRE ET DE GROUPE SONT REPORTEES LES VALEURS DE SUPPORTS AVEC ENTRE PARENTHESE LE NOMBRE D'ESPECES ETUDIEE.

	Manos et al, 1999 (Figure A17)		Manos and Standford 2001 (Figure A18)	Manos et al, 2001 (Figure A13)	Bellarosa et al, 2005 (Figure A19)	Oh and Manos, 2008 (Figure A15)	Denk and Grimm, 2010 (Figure A14)
	ITS, parcimonie	ITS + restriction, parcimonie	ITS, parcimonie	ITS, parcimonie	ITS, parcimonie	CRC+ITS, ML	ITS, ML
Taille du fragment	614 pb	(614 + 700)	643 pb	1038 pb	600 pb	3624 pb	854 pb
Fagus	(0)	(0)	(0)	(1)	(0)	100 (2)	(0)
Colombalanus	(1)	(1)	(0)	(1)	(1)	(1)	(1)
Trigonobalanus	(1)	(1)	(0)	(1)	(1)	(1)	(1)
Formanodendron	0)	(0)	(0)	72 (2)	(0)	(1)	(1)
Lithocarpus	(0)	(0)	(1)	93 (61)	(0)	99 (9)	87 (48)
Chrysolepis	(0)	(0)	(1)	88 (2)	(0)	100 (2)	96 (2)
Notholithocarpus	(0)	(0)	0)	100 (2)	(0)	100 (2)	96 (2)
Castanea	(0)	(0)	(1)	91 (4)	(0)	100 (2)	93 (4)
Castanopsis	(0)	(0)	(1)	52 (22)	(0)	100 (4)	< 50 (22)
Cyclobalanopsis	(1)	(1)	97 (7)	62 (9)	(0)	100 (3)	65 (8)
<i>Cerris</i>	99 (3)	87 (4)	78 (14)	92 (20)	(1)	100 (2)	88 (11)
<i>Ilex</i>	<50 (4)	65 (3)		<50 (7)	100 (2)	(1)	30 (16)
Protobalanus	65 (4)	78 (5)	100 (6)	60 (6)	(0)	(1)	(5)
<i>Quercus s.s</i>	<50 (10)	90 (8)	Polytomie (10+4)	Polytomie (20+6)	99 (5)	96 (6)	(41)
Lobatae	100 (4)	100 (4)			(0)	100 (4)	(13)

Il est également intéressant de noter que les marqueurs chloroplastiques utilisés avec succès classiquement en phylogénie chez les autres genres de plantes (MatK, rBcl) ainsi que les marqueurs nucléaires (ITS) sont tous caractérisés dans la famille des Fagacées par un manque de résolution sur les nœuds profonds, suggérant ainsi qu'un phénomène biologique ait touché ces deux génomes de la même façon. Il est possible que la (ou les) radiations ait été si rapide que les gènes classiques ne soient donc pas de bons marqueurs. En effet, à taux d'évolution constant au cours du temps, pour qu'une radiation rapide soit détectée, il est nécessaire que le marqueur moléculaire interrogé mute rapidement, le problème étant que si en plus la radiation est ancienne, ce marqueur risque d'être saturé dans le temps et que l'information soit perdue. Une autre hypothèse expliquant le manque résolution aux nœuds profonds a été suggérée par Denk (Denk *et al.*, 2012). En effet, il est envisageable que l'évolution se soit déroulée sous la forme de réticulations (hybridation entre espèces, l'hybride devenant lui-même une espèce à part entière par la suite) puis il y'a eu extinctions de certaines espèces conduisant à une irrésolution car la phylogénie moléculaire ne se fait que sur les espèces actuelles.

Notre hypothèse est que le manque de soutien des branches reliées aux nœuds profonds de la phylogénie la plus résolutive à ce jour (CRC+ ITS) conduisant à l'obtention des deux groupes « Old World » et « New World Oaks » ne soit qu'un symptôme d'une irrésolution artefactuelle du nœud plus profond situé entre la racine et ces deux groupes ; et qu'en fait, ces deux groupes ne sont pas indépendants mais dérivent l'un de l'autre.

Il semble donc nécessaire si cette hypothèse est vérifiée, qu'il faille plutôt rechercher des marqueurs mutant lentement mais en nombre important. En effet, un gène mutant lentement ne donnera pas ou peu d'information phylogénétique seul, mais, combiné avec un nombre important d'autres marqueurs, le signal phylogénétique peut s'amplifier et donner le signal adéquat.

Le travail de la présente thèse se fait dans cet esprit : à travers un échantillonnage plus complet que les études précédentes et la recherche de nouveaux gènes, nous espérons trouver une combinaison de gènes capable de résoudre la phylogénie du genre *Quercus* au niveau des nœuds profonds (si cela existe) avec des valeurs de soutien de

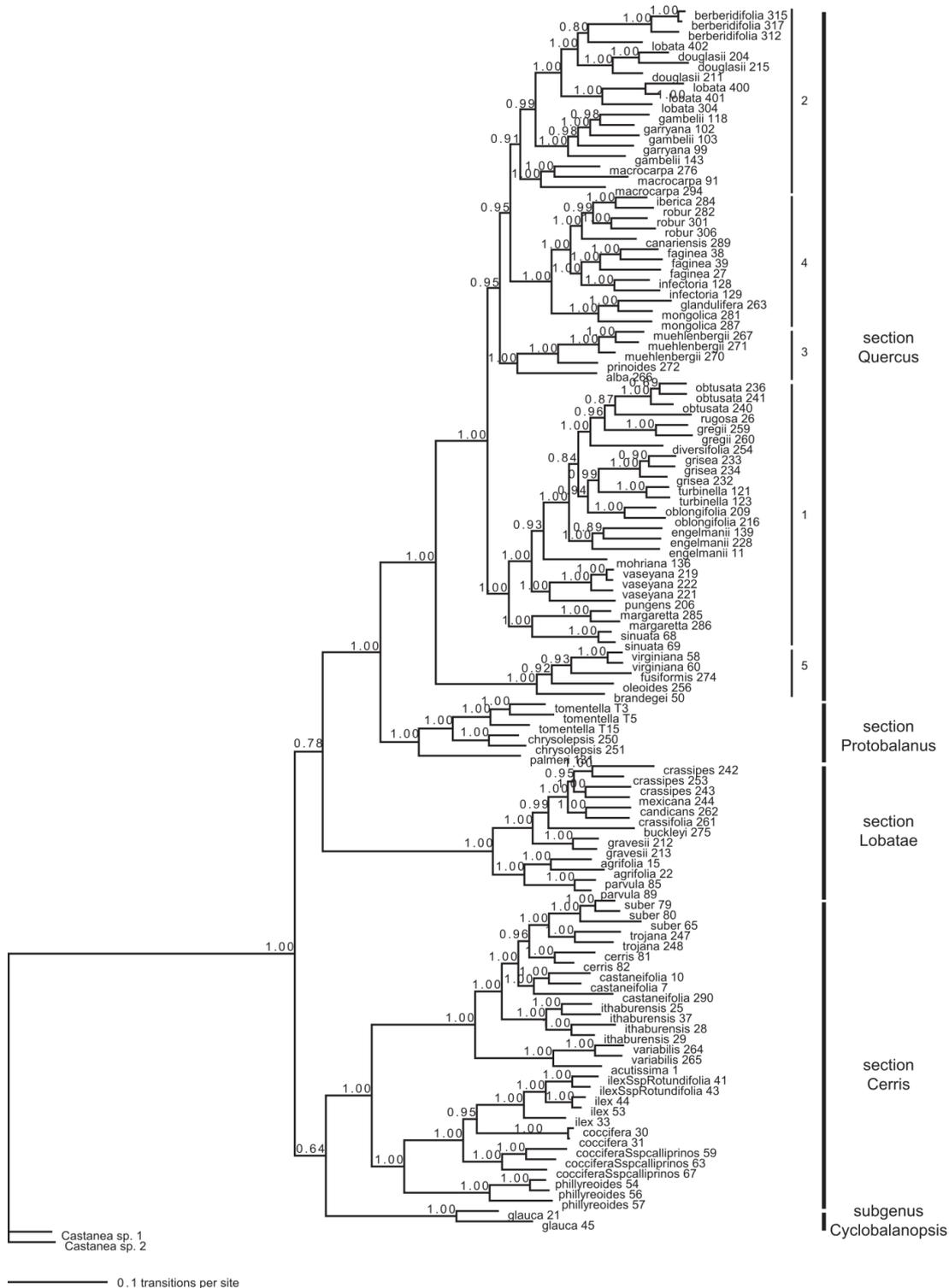


FIGURE A19 : ARBRE DE CONSENSUS MAJORITAIRE ISSU DE DEUX RUNS INDEPENDANTS DE 1 MILLION DE GENERATION, BASE SUR DES DONNEES DE TYPE AFLP (PEARSE AND HIPPI, 2009).

branches élevés, ou bien, si les résultats donnent une irrésolution comparable aux études précédentes, nous pencherons pour l'hypothèse selon laquelle la phylogénie moléculaire nucléaire de ce genre est impossible en l'état à cause du fait que nous n'avons pas accès aux espèces éteintes et/ou que les marqueurs utilisés ne sont pas adéquats pour répondre à cette interrogation.

L'exploration de marqueurs moléculaires s'est faite sur deux types différents : des marqueurs nucléaires choisis grâce aux ressources génomiques dont dispose le laboratoire d'accueil sur 110 espèces, le but étant de voir si la phylogénie nucléaire obtenue bute sur les mêmes problèmes ou non que les études précédentes (irrésolution due aux nœuds profonds). En outre, sur un sous échantillon de ces espèces, le séquençage total du génome chloroplastique a été réalisé et la comparaison des phylogénies de ces deux génomes peut aider dans la compréhension des problèmes rencontrés.



### A3.3 REFERENCES BIBLIOGRAPHIQUES

---

- Bellarosa, R., M. C. Simeone, et al. (2005). Utility of ITS sequence data for phylogenetic reconstruction of Italian *Quercus* spp. *Molecular Phylogenetics and Evolution* **34**(2): 355-370.
- Bremer, B., K. Bremer, et al. (2009). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society* **161**(2): 105-121.
- Camus, A. (1936-1954). Les chênes, Monographie du genre *Quercus* et Monographie du genre *Lithocarpus*.
- Denk, T. and G. W. Grimm (2010). The oaks of western Eurasia: Traditional classifications and evidence from two nuclear markers. *Taxon* **59**(2): 351-366.
- Forman, L. L. (1966). On the evolution of cupules in the Fagaceae. *Kew Bulletin* **18**(3): 385-419.
- Jones, J. H. (1986). Evolution of the Fagaceae - the Implications of Foliar Features. *Annals of the Missouri Botanical Garden* **73**(2): 228-275.
- Manos, P. S., C. H. Cannon, et al. (2008). Phylogenetic Relationships and Taxonomic Status Of the Paleoendemic Fagaceae Of Western North America: Recognition Of A New Genus, *Notholithocarpus*. *Madroño* **55**(3): 181-190.
- Manos, P. S., J. J. Doyle, et al. (1999). Phylogeny, Biogeography, and Processes of Molecular Differentiation in *Quercus* Subgenus *Quercus* (Fagaceae). *Molecular Phylogenetics and Evolution* **12**(3): 333-349.
- Manos, P. S. and A. M. Stanford (2001). The Historical Biogeography of Fagaceae: Tracking the Tertiary History of Temperate and Subtropical Forests of the Northern Hemisphere. *International Journal of Plant Sciences* **162**(S6): S77-S93.
- Manos, P. S., Z. K. Zhou, et al. (2001). Systematics of Fagaceae: Phylogenetic Tests of Reproductive Trait Evolution. *International Journal of Plant Sciences* **162**(6): 1361-1379.
- Menitsky, J. L. (1984). *Duby Asii*. Leningrad: Nauka. .
- Menitsky, Y. L. (2005). *Oaks of Asia*, Science Publishers.



- Nixon, K. (1985). A biosystematic study of *Quercus* section *Virentes* (the live oaks) with phylogenetic analyses of Fagales, Fagaceae and *Quercus*. Austin, University of Texas.
- Nixon, K. (1993). Infrageneric classification of *Quercus* (Fagaceae) and typification of sectional names. Ann. For. Sci. **50**(Supplement): 25s-34s.
- Nixon, K. C. (1997). Fagaceae, Oxford University Press, New York.
- Nixon, K. C. and W. L. Crepet (1989). *Trigonobalanus* (Fagaceae) - Taxonomic Status and Phylogenetic-Relationships. American Journal of Botany **76**(6): 828-841.
- Oh, S. H. and P. S. Manos (2008). Molecular phylogenetics and cupule evolution in Fagaceae as inferred from nuclear CRABS CLAW sequences. Taxon **57**(2): 434-451.
- Ohba, H. (1989). Fagaceae. Wild Flowers of Japan. Y. Satake, H. Hara, S. Watari and T. Tominari. Tokyo. **1**: 66-78.
- Pearse, I. S. and A. L. Hipp (2009). Phylogenetic and trait similarity to a native species predict herbivory on non-native oaks. Proceedings of the National Academy of Sciences **106**(43): 18097-18102.
- Samuel, R., A. Bachmair, et al. (1998). ITS sequences from nuclear rDNA suggest unexpected phylogenetic relationships between Euro-Mediterranean, East Asiatic and North American taxa of *Quercus* (Fagaceae). Plant Systematics and Evolution **211**(1-2): 129-139.
- Schwarz, O. I., T. G. Tutin, et al. (1964). *Quercus* L. Flora Europaea. Cambridge, Cambridge University Press. **1 : Lycopodiaceae to Platanaceae**.
- Soepadmo, E. (1972). Fagaceae. Flora Malenesia.
- Valencia, S. A. (2004). Diversidad del genero *Quercus* (Fagaceae) en Mexico. . Bol. Soc. Bot. Mex. **75**: 33-54.
- Xu, L. (2004). Diversité de l'ADN chloroplastique et relations phylogénétiques au sein des Fagacées et du genre *Quercus*. Thèse de l'Université Henri Poincaré, Nancy, France.
- Zheng, W. (1985). Chinese Arboretum. Beijing, Chinese Forestry press.

TABLEAU A6 : QUELQUES CALIBRATIONS FOSSILES UTILISEES (DAGHLIAN AND CREPET, 1983; DENK ET AL., 2012).

	<b>Groupe</b>	<b>Type de fossile</b>	<b>Pays</b>	<b>Datation</b>
<i>Quercus oligocenensis</i>	<i>Lobatae</i>	Fruit	USA	28.4-33.9
<i>Quercus huntsvillensis</i>	<i>Quercus s.s</i>	Fleurs, feuilles	USA	28.4-33.9
<i>Quercus ilex</i>	<i>Ilex</i>	Pollen	Allemagne	28.4-33.9
<i>Quercus hannibili</i>	<i>Protobalanus</i>	Feuilles, fruits	USA	15.3

TABLEAU A7 : DATATIONS FOSSILES DU GROUPE DES CYCLOBALANOPSIS. LA DATATION UTILISEE EST EN GRAS.

LE NUMERO INDIQUE DANS LA DERNIERE COLONNE CORRESPOND A LA REFERENCE DE L'ECHANTILLON DANS LA BASE DE DONNEES : [HTTP://WWW.PALEODB.ORG/CGI-BIN/BRIDGE.PL](http://www.paleodb.org/cgi-bin/bridge.pl).

<b>Pays</b>	<b>Age début (ma)</b>	<b>Age fin (ma)</b>	<b>Espèce fossile</b>	<b>Type de fossile</b>	<b>Numéros base de données/article</b>
Japon	<b>55.8</b>	<b>33.9</b>	Cyclobalanopsis naitoi	Feuille	24465
Japon	<b>55.8</b>	<b>33.9</b>	Cyclobalanopsis naitoi	Feuille	24466
Japon	<b>55.8</b>	<b>33.9</b>	Cyclobalanopsis naitoi	Feuille	24474
Japon	<b>55.8</b>	<b>33.9</b>	Cyclobalanopsis naitoi	Feuille	24488
Japon	<b>55.8</b>	<b>33.9</b>	Cyclobalanopsis naitoi	Feuille	24490

## A4 BIOGEOGRAPHIE HISTORIQUE DU GENRE *QUERCUS*

---

### A4.1 EVOLUTION DU GENRE *QUERCUS*

---

#### A4.1.1 DONNEES FOSSILES

---

L'histoire de la famille des Fagacées peut se voir grâce à la richesse spécifique qui est enregistrée dans les sédiments de l'âge tertiaire (-65 millions d'années à -1.8 millions d'années) en Amérique et en Eurasie (Crepet and Daghljan, 1980; Daghljan and Crepet, 1983; Crepet and Nixon, 1989; Manos and Stanford, 2001 ). L'âge minimum de l'apparition de la famille se situe aux alentours des 53 millions d'années (Crepet and Nixon, 1989; Nixon and Crepet, 1989; Manos and Stanford, 2001a). Des macrofossiles de types *Castaneoides*, *Trigonobalanoides*, *Quercoides* et *Fagoides* identifiés sans équivoque ont été datés autour de cette période, indiquant une rapide diversification au sein de la famille (Crepet, 1989; Meyer and Manchester, 1997; Manos and Stanford, 2001 ). Des restes fossiles de feuilles attribués au genre *Quercus* et ressemblant fortement aux espèces actuelles indiquent que la diversification infragénérique était déjà établie entre -34 et -23 millions d'années (Axelrod, 1983; Crepet, 1989; Manos and Stanford, 2001 ). Axelrod (1983) suggéra à travers la biogéographie du genre *Quercus* et de son histoire riche en fossiles, que la majorité des groupes ont évolués in situ avec des radiations principalement locales et quelques groupes monophylétiques ayant une distribution transcontinentale. Les données fossiles suggèrent une origine nord américaine plus ancienne du genre durant l'Oligocène (-34 à -23 millions d'années) et une apparition plus tardive en Chine après -23 millions d'années (Daghljan and Crepet, 1983; Zhou, 1992, 1993; Manos and Stanford, 2001 ).

TABLEAU A8 : DATATIONS FOSSILES DU GROUPE DES CERRIS. LA DATATION UTILISEE EST EN GRAS.

LE NUMERO INDIQUE DANS LA DERNIERE COLONNE CORRESPOND A LA REFERENCE DE L'ECHANTILLON DANS LA BASE DE DONNEES : [HTTP://WWW.PALEODB.ORG/CGI-BIN/BRIDGE.PL](http://www.paleodb.org/cgi-bin/bridge.pl).

Pays	Age début (ma)	Age fin (ma)	Espèce fossile	Type de fossile	Numéros base de données/article
Géorgie	11.6	3.6	<i>Quercus kubinyii (B)</i>	Graine	(1)
République Tchèque	23.0	15.9	<i>Quercus kubinyii (B)</i>	Non disponible	(2)
République Tchèque	23.0	15.9	<i>Quercus kubinyii (B)</i>	Non disponible	(2)
<b>France</b>	<b>40.4</b>	<b>37.2</b>	<b><i>Quercus kubinyii (B)</i></b>	<b>Non disponible</b>	<b>(1)</b>
France	5.3	3.6	<i>Quercus kubinyii (B)</i>	Non disponible	(1)
France	5.3	3.6	<i>Quercus kubinyii (B)</i>	Non disponible	(1)
Allemagne	13.7	11.6	<i>Quercus kubinyii (B)</i>	Feuilles	27512
Allemagne	5.3	1.8	<i>Quercus trojana</i>	Graine	28171
Allemagne	5.3	3.6	<i>Quercus trojana</i>	Graine	28134
Grèce	7.2	5.3	<i>Quercus kubinyii (B)</i>	Graine	(1)
Italie	7.2	5.3	<i>Quercus gigas (A)</i>	Non disponible	(1)
Italie	7.2	5.3	<i>Quercus gigas (A)</i>	Non disponible	(1)
Italie	7.2	5.3	<i>Quercus gigas (A)</i>	Non disponible	(1)
Italie	7.2	5.3	<i>Quercus gigas (A)</i>	Non disponible	(1)

Pologne	16.0	11.6	<i>Quercus kubinyii (B)</i>	Feuilles	28442
Japon	23.0	1.8	<i>Quercus miovariabilis (C)</i>	Non disponible	22786
Japon	11.6	5.3	<i>Quercus miovariabilis (C)</i>	Non disponible	21239
Japon	11.6	5.3	<i>Quercus miovariabilis (C)</i>	Non disponible	21256
Japon	11.6	5.3	<i>Quercus miovariabilis (C)</i>	Non disponible	21262
Japon	11.6	5.3	<i>Quercus miovariabilis (C)</i>	Non disponible	21234
Japon	16.0	5.3	<i>Quercus miovariabilis (C)</i>	Non disponible	22739
Japon	15.9	11.6	<i>Quercus miovariabilis (C)</i>	Bois	(3)
Japon	23.0	1.8	<i>Quercus miovariabilis (C)</i>	Non disponible	22796
Japon	65.5	0.0	<i>Quercus miovariabilis (C)</i>	Feuilles	22768
Japon	15.9	11.6	<i>Quercus miovariabilis (C)</i>	Feuilles	(3)
Japon	65.5	0.0	<i>Quercus sinomiocenicum (C)</i>	Non disponible	22826

(A) : [http://www.bio.bas.bg/~phytolbalcan/PDF/17\\_1/17\\_1\\_01\\_Bozukov\\_&\\_al.pdf](http://www.bio.bas.bg/~phytolbalcan/PDF/17_1/17_1_01_Bozukov_&_al.pdf)

(B) : <http://biostor.org/cache/pdf/b9/b1/4c/b9b14c3777da09d710744f77a4097283.pdf>

(C) : [http://www.ib-pan.krakow.pl/pubs-pdf/Acta%20Palaeobotanica/2000/Ac402\\_s153-163.pdf](http://www.ib-pan.krakow.pl/pubs-pdf/Acta%20Palaeobotanica/2000/Ac402_s153-163.pdf)

(1) : Kovar-Eacekder, J., Kvacek, Martinetto, ., Roiron, P., 2006, Late Miocene to Early Pliocene vegetation of southern Europe (7-4Ma) as reflected in the megafossil plant record, *Palaeogeography, Palaeoclimatology, Palaeoecology* 238 (2006) 321-339.

(2) : KVAČEK, Z. & TEODORIDIS, V. 2007. Tertiary macrofloras of the Bohemian Massif: a review with correlations within Boreal and Central Europe. *Bulletin of Geosciences* 82(4), 383-408 (5 figures, 12 tables). Czech Geological Survey, Prague. ISSN 1214-1119. Manuscript received August 15, 2007, accepted in revised form October 4, 2007, issued December.

(3) : Yabe A. 2008. Early Miocene terrestrial climate inferred from plant megafossil assemblages of the Joban and Soma areas, Northeast Honshu, Japan. *Bulletin of the Geological Survey of Japan*, vol.59 (7/8), pp. 397-413.



---

#### A4.1.2 SCENARIOS EVOLUTIFS PRINCIPAUX EVOQUES

---

Les données fossiles ne permettent pas de trancher en faveur d'une hypothèse évolutive unique. Ainsi, il existe deux principaux scénarios : celui de Zhou (1992) qui suggère une origine du genre *Quercus* en Asie du Sud-Est. *Quercus* aurait dérivé du genre *Trigonobalanus* dans les montagnes tropicales de l'Indochine aux alentours de -65 millions d'années et aurait ensuite migré dans deux directions :

\* Avant -53 millions d'années vers l'Europe et via le NALB (North Atlantic Land Bridge) vers l'Amérique du Nord ; en effet, les connections entre l'Europe et l'Amérique du Nord étaient possibles à cette époque, le NALB se fermant il y a 23 millions d'années. Le groupe *Cerris* aurait divergé à ce moment en donnant les groupes *Ilex* et les *Protobalanus*.

\* Après 23 millions d'années, en empruntant le détroit de Béring vers l'Amérique du Nord, ce sont les *Quercus s.s* qui auraient colonisé le Nouveau-Monde à ce moment là.

Le scénario de Zhou (Zhou, 1993) est essentiellement basé sur des données biogéographiques actuelles et passées.

L'hypothèse alternative d'évolution du genre provient de l'analyse faite par l'équipe de Manos (2001) sur les données ITS vues précédemment, à l'aide d'un logiciel de biogéographie. L'importance de la résolution de l'arbre phylogénétique produit prend toute son importance à ce stade.

Pour rappel, cette étude avait conclu sur l'existence de deux grands groupes (« New World Oaks » et « Old World Oaks ») : un groupe contenant les espèces eurasiatiques (*Cerris+Ilex+Cyclobalanopsis*) et un groupe contenant majoritairement des espèces américaines (*Protobalanus+Quercus s.s + Lobatae*).

Les *Quercus s.s* font figure d'exception car c'est l'unique groupe qui a des représentants dans les deux continents.

TABLEAU A9 : DATATIONS FOSSILES DU GROUPE DES ILEX. LA DATATION UTILISEE EST EN GRAS.

LE NUMERO INDIQUE DANS LA DERNIERE COLONNE CORRESPOND A LA REFERENCE DE L'ECHANTILLON DANS LA BASE DE DONNEES : [HTTP://WWW.PALEODDB.ORG/CGI-BIN/BRIDGE.PL](http://www.paleodb.org/cgi-bin/bridge.pl).

Pays	Age début (ma)	Age fin (ma)	Espèce fossile	Type de fossile	Numéros base de données/article
Bulgarie	7.246	5.332	<i>Quercus mediterranea (A)</i>	Graine	(1)
Crête	23.03	5.332	<i>Quercus mediterranea (A)</i>	Feuilles	(2)
France	5.332	3.6	<i>Quercus coccifera</i>	Non disponible	(1)
France	1.8	0.0	<i>Quercus ilex</i>	Pollen	50522
France	5.332	3.6	<i>Quercus ilex</i>	Non disponible	(1)
France	5.332	3.6	<i>Quercus ilex</i>	Non disponible	(1)
France	1.8	0.8	<i>Quercus type ilex-coccifera</i>	Non disponible	50342
<b>France/Espagne</b>	<b>40.4</b>	<b>37.2</b>	<b><i>Quercus mediterranea (A)</i></b>	<b>Non disponible</b>	<b>(1)</b>
Grèce	7.246	5.332	<i>Quercus mediterranea (A)</i>	Graine	(1)
Grèce	7.246	5.332	<i>Quercus mediterranea (A)</i>	Graine	(1)
Italie	7.246	5.332	<i>Quercus mediterranea (A)</i>	Non disponible	(1)
Italie	7.246	5.332	<i>Quercus mediterranea (A)</i>	Non disponible	(1)
Italie	7.246	5.332	<i>Quercus mediterranea (A)</i>	Non disponible	(1)
Maroc	11.608	5.332	<i>Quercus type-ilex-coccifera</i>	Pollen	(3)
Maroc	11.608	5.332	<i>Quercus type-ilex-coccifera</i>	Pollen	(3)

(A) : <http://biostor.org/cache/pdf/b9/b1/4c/b9b14c3777da09d710744f77a4097283.pdf>

(1) : Kovar-Eacekder, J., Kvacek, Martinetto, ., Roiron, P., 2006, Late Miocene to Early Pliocene vegetation of southern Europe (7-4Ma) as reflected in the megafossil plant record, Palaeogeography, Palaeoclimatology, Palaeoecology 238 (2006) 321-339.

(2) : ZIDIANAKIS G., MOHR B. A. R. and FASSOULAS C. 2007. A late Miocene leaf assemblage from Vrysses, western Crete, Greece, and its paleoenvironmental and paleoclimatic interpretation. GEODIVERSITAS, 29 (3), pp. 351-377.

(3) : BACHIRI TAOUFIQ N., BARHOUN N. and SUC J.-P. 2008. Les environnements continentaux du corridorifain (Maroc) au Miocène supérieur d'après la palynologie. Geodiversitas 30 (1), pp. 41-58.

Le résultat de cette étude (Manos and Stanford, 2001 ) rejoint l'hypothèse de Trelease (1924) selon laquelle, un groupe ancestral aurait eu une distribution transcontinentale très vaste, et à la suite d'isolements dus aux fermetures des voies de communications, le phénomène de vicariance aurait donné ces deux grands groupes. Ce groupe ancestral aurait donné naissance :

\* En Asie : aux *Cyclobalanopsis* et à la section *Cerris* (ces auteurs ne différencient pas le groupe *Ilex* du groupe *Cerris*) ; les espèces de ce dernier groupe auraient migré vers l'Europe en effectuant de nombreux allers-retours.

\* En Amérique du Nord : aux sections *Quercus s.s* et *Lobatae*. La section *Quercus s.s.* a migré vers le continent Asiatique puis Européen via le détroit de Béring. Ce scénario est en accord avec le plus faible nombre d'espèces de chênes blancs (*Quercus s.s*) en Eurasie qu'en Amérique et l'apparition plus ancienne de fossiles en Amérique (Zhou, 1992, 1993; Manos and Stanford, 2001 ). De plus, l'estimation du temps de divergence entre les *Quercus s.s* Américains et Eurasiatiques étant de -17 millions d'années (calcul fait à partir d'une estimation de l'horloge moléculaire du gène *rbcl*), il est en concordance avec la possibilité de migration via Béring qui était effective après -23 millions d'années (Manos and Stanford, 2001 ).

Une remarque importante doit d'être faite à ce stade. En effet, les explications concernant les hypothèses sur l'origine du genre *Quercus* font principalement appel à des données biologiques anciennes (fossiles), de génétique (phylogénies moléculaires) ou des données géologiques (routes de migrations) mais de façon très anecdotique aux données paléoclimatiques.

Ce dernier type de données peut donner des informations primordiales et complémentaires aux données géologiques, dans le sens où, connaissant les exigences actuelles de ce genre et en partant du principe qu'elles furent les mêmes dans le passé, la connaissance du climat à certaines périodes données et en certains lieux bien précis, peut indiquer les routes empruntées lors des migrations passées. Ce travail fait partie intégrante d'une collaboration entre l'équipe à laquelle j'appartiens (BIOGECO) et le LSCE (Laboratoire des Sciences du Climat et de l'Environnement) dans le cadre du projet PhyloSpace.

TABLEAU A10 : DATATIONS FOSSILES DES *QUERCUS* S.S. LA DATATION UTILISEE EST EN GRAS. LE NUMERO INDIQUE DANS LA DERNIERE COLONNE CORRESPOND A LA REFERENCE DE L'ECHANTILLON DANS LA BASE DE DONNEES : [HTTP://WWW.PALEODB.ORG/CGI-BIN/BRIDGE.PL](http://www.paleodb.org/cgi-bin/bridge.pl).

Pays	Age début (ma)	Age fin (ma)	Espèce fossile	Type de fossile	Numéros base de données/article
Géorgie	11.608	3.6	<i>Quercus petraea</i> -type	Graine	(1)
<b>Russie</b>	<b>55.8</b>	<b>40.4</b>	<b><i>Quercus prinus</i></b>	<b>Pollen</b>	<b>(2)</b>
France	5.332	3.6	<i>Quercus faginea</i>	Non disponible	(1)
France	0.8	0.1	<i>Quercus frainetto</i>	Non disponible	42673
France	0.8	0.1	<i>Quercus lusitanica</i>	Non disponible	42673
France	1.8	0.8	<i>Quercus robur</i>	Non disponible	50524
France	1.8	0.8	<i>Quercus pendunculata</i> -pubescens type	Non disponible	50342
Allemagne	3.6	2.6	<i>Quercus pubescens</i>	Graine	22778
Allemagne	11.6	5.3	<i>Quercus robur</i>	Graine	27921
Allemagne	3.6		<i>Quercus robur</i>	Graine	28197
Allemagne	3.6	2.6	<i>Quercus robur</i>	Graine	28198
Allemagne	5.3	3.6	<i>Quercus robur</i>	Graine	28134
Allemagne	13.7	11.6	<i>Quercus roburoides</i>	Feuilles	27377
Pologne	16.0	11.6	<i>Quercus pubescens</i>	Graine	28442

USA	1.8	0.0	<i>Quercus virginiana</i>	Non disponible	34269
USA	10.5	8.5	<i>Quercus columbiana</i>	Feuilles	(3)
USA	23.03	5.332	<i>Quercus columbiana</i>	Feuilles	(3)
USA	23.03	5.332	<i>Quercus columbiana</i>	Feuilles	(3)
USA	10.3	4.9	<i>Quercus douglasoides</i>	Feuilles	35271
USA	5.3	1.8	<i>Quercus douglasoides</i>	Feuilles	35272
USA	37.2	28.4	<i>Quercus dumosoides</i>	Feuilles	39837
USA	37.2	28.4	<i>Quercus dumosoides</i>	Feuilles	39839
USA	65.5	0.0	<i>Quercus dumosoides</i>	Non disponible	24209
USA	46.2	42.0	<i>Quercus petros</i>	Non disponible	2139
USA	10.5	8.5	<i>Quercus prelobata</i>	Feuilles	(3)
USA	23.03	5.332	<i>Quercus prelobata</i>	Feuilles	(3)
USA	23.03	5.332	<i>Quercus prelobata</i>	Feuilles	(3)
USA	5.3	1.8	<i>Quercus prelobata</i>	Feuilles	35908
USA	23.0	5.3	<i>Quercus prelobata</i>	Non disponible	23973
USA	23.0	5.3	<i>Quercus treleasii</i>	Feuilles	103753
Japon	11.6	5.3	<i>Quercus miocripula(A)</i>	Non disponible	22755
Japon	23.0	1.8	<i>Quercus miocripula (A)</i>	Non disponible	22796
Japon	65.5	0.0	<i>Quercus protoaliena</i>	Non disponible	22819

Japon	11.6	5.4	<i>Quercus protoaliena</i>	Non disponible	21270
Japon	16.0	5.3	<i>Quercus protoaliena</i>	Non disponible	22744
Japon	23.0	1.8	<i>Quercus protoaliena</i>	Non disponible	22796
Japon	65.5	0.0	<i>Quercus protoaliena</i>	Non disponible	22826

(A) :

[http://159.226.74.248:8000/searchbycontentnew.jsp?showdetail=1&wztm=A+Revision+of+%3Ci%3EQuercus+miocripula%3C%2Fi%3E+Huzioka+\(Subsect.+Diversipilosae%2C+Sect.+Lepidobalanus\)+from+the+Miocene+of+Eastern+Asia](http://159.226.74.248:8000/searchbycontentnew.jsp?showdetail=1&wztm=A+Revision+of+%3Ci%3EQuercus+miocripula%3C%2Fi%3E+Huzioka+(Subsect.+Diversipilosae%2C+Sect.+Lepidobalanus)+from+the+Miocene+of+Eastern+Asia)

(1) : Kovar-Eeckder, J., Kvacek, Martinetto, ., Roiron, P., 2006, Late Miocene to Early Pliocene vegetation of southern Europe (7-4Ma) as reflected in the megafossil plant record, *Palaeogeography, Palaeoclimatology, Palaeoecology* 238 (2006) 321-339.

(2) : A.A. Velichko, E.F. Akhlestina, O.K. Borisova, Yu.N. Gribchnko, N.Ya. Zhidovinov, E.M. Zelikson, Yu.I. Iosifova, V.A. Klimanov, T.D. Morosova, V.P. Nechaev, V.V. Pisareva, T.V. Svetlitskaya, I.I. Spasskaya, V.P. Udartsev, M.A. Faustova and S.M. Shik. 2005. East European Plain in Cenozoic climatic and environmental changes in Russia. A. A. Velichko and V. P. Nechaev. *The geological society of America*, 2005, pp. 226.

(3) : BUECHLER W. K., MICHAEL T. DUNN, AND WILLIAM C. REMBER, 2007. LATE MIOCENE PICKETT CREEK FLORA OF OWYHEE COUNTY, IDAHO, *CONTRIBUTIONS FROM THE MUSEUM OF PALEONTOLOGY THE UNIVERSITY OF MICHIGAN*, VOL. 31, NO. 12, PP. 305-362.

L'objectif de la présente thèse étant de fournir une phylogénie du genre *Quercus* calibrée à l'aide de données fossiles, ce travail sera continué et mis en relation avec les cartes paléoclimatiques produites. C'est pourquoi la partie suivante ne traitera pas des phénomènes climatiques qui se sont déroulés dans le passé, mais se concentrera (si elle est réalisable) sur l'obtention de calibration d'une échelle de temps sur les reconstructions phylogénétiques s'inspirant des données fossiles disponibles. De plus, un ou plusieurs scénarios de biogéographie historique seront proposés à la suite de l'analyse des phylogénies moléculaires produites.

#### A4.2 APPORTS DE LA BASE DE DONNEES FOSSIL MAP A LA CALIBRATION

---

Une revue bibliographique a été réalisée dans le cadre du projet PhyloSPACE par deux étudiantes : Ludivine Chabaud et Aurélie Ajas. Le but étant de compiler toute l'information disponible possible sur les fossiles du tertiaire (-65 ma à -1.8 ma). Cette base de données (FossilMap) est accessible sur le site : <http://w3.pierroton.inra.fr/fossile/index.php?p=fmap>, et comprend les fossiles décrits dans la littérature sous forme d'articles, mais également répertoriés dans des bases de données paléontologiques comme paleodb (<http://www.paleodb.org/cgi-bin/bridge.pl>). FossilMap regroupe toutes les données disponibles sur les Fagacées, et en particulier plus de 350 références de *Quercus*.

Il faut signaler qu'il est très difficile de juger a posteriori de la qualité des données issues de la littérature du point de vue de l'attribution à une espèce mais également de la datation utilisée.

Afin d'essayer de limiter au mieux ce problème, une recherche sur internet a été faite à la suite de ce travail afin de voir si les fossiles retrouvés avaient ou non une affinité avec un des six groupes actuels (tels que définis au chapitre A1-2), seuls les fossiles de ce type seront discutés et inclus dans le texte, les autres seront éliminés. Encore une fois, nous sommes obligés de faire confiance aux résultats de la littérature, car nous n'avons pas de compétences en paléobotanique.

Il n'existe pas dans la littérature, de phylogénie calibrée par des datations fossiles sur le genre *Quercus*. L'équipe de Manos qui comme nous l'avons vu précédemment a beaucoup travaillé sur le genre, utilise quatre âges de référence (communications personnelles), correspondant à des fossiles bien datés (Tableau A6).

TABLEAU A11 : DATATIONS FOSSILES DES *PROTOBALANUS*. LA DATATION UTILISEE EST EN GRAS.

LE NUMERO INDIQUE DANS LA DERNIERE COLONNE CORRESPOND A LA REFERENCE DE L'ECHANTILLON DANS LA BASE DE DONNEES : [HTTP://WWW.PALEODB.ORG/CGI-BIN/BRIDGE.PL](http://www.paleodb.org/cgi-bin/bridge.pl).

Pays	Age début (ma)	Age fin (ma)	Espèce fossile	Type de fossile	Numéros base de données/article
USA	10.5	8.5	<i>Quercus hannibali</i> (A)	Feuilles	(1)
USA	13.6	10.3	<i>Quercus hannibali</i> (A)	Feuilles	27659
USA	13.6	10.3	<i>Quercus hannibali</i> (A)	Feuilles	27661
a	13.6	10.3	<i>Quercus hannibali</i> (A)	Feuilles	27662
USA	16.0	11.6	<i>Quercus hannibali</i> (A)	Feuilles	35412
USA	16.3	13.6	<i>Quercus hannibali</i> (A)	Feuilles	27660
USA	23.0	5.3	<i>Quercus hannibali</i> (A)	Feuilles	35278
USA	23.03	5.332	<i>Quercus hannibali</i> (A)	Feuilles	(2)
USA	33.9	5.3	<i>Quercus hannibali</i> (A)	Non disponible	23973
USA	33.9	5.3	<i>Quercus hannibali</i> (A)	Non disponible	23972
USA	33.9	5.33	<i>Quercus hannibali</i> (A)	Non disponible	23975
USA	20.4	16.0	<i>Quercus hannibali</i> (A)	Plusieurs organes	27702
USA	11.6	5.3	<i>Quercus hannibali</i> (A)	Feuilles	35270
USA	20.4	16.0	<i>Quercus hannibali</i> (A)	Plusieurs organes	27701
USA	20.4	16.0	<i>Quercus hannibali</i> (A)	Plusieurs organes	27179
<b>USA</b>	<b>33.9</b>	<b>28.4</b>	<b><i>Quercus moyei</i> (B)</b>	<b>Feuilles</b>	<b>27569</b>
USA	5.3	1.8	<i>Quercus pliopalmerii</i>	Feuilles	35272

(A) + (1) : Daghlian CP, Crepet WL. Oak catkins, leaves and fruits from the Oligocene Catahoula Formation and their evolutionary significance. Am J Bot (1983) 70:639-649.

(2) : BUECHLER W. K., MICHAEL T. DUNN, AND WILLIAM C. REMBER, 2007. LATE MIOCENE PICKETT CREEK FLORA OF OWYHEE COUNTY, IDAHO, CONTRIBUTIONS FROM THE MUSEUM OF PALEONTOLOGY THE UNIVERSITY OF MICHIGAN, VOL. 31, NO. 12, PP. 305-362.

TABLEAU A12 : DATATIONS FOSSILES DU GROUPE DES *LOBATAE*. LA DATATION UTILISEE EST EN GRAS.

LE NUMERO INDIQUE DANS LA DERNIERE COLONNE CORRESPOND A LA REFERENCE DE L'ECHANTILLON DANS LA BASE DE DONNEES : [HTTP://WWW.PALEODB.ORG/CGI-BIN/BRIDGE.PL](http://www.paleodb.org/cgi-bin/bridge.pl).

Pays	Age début	Age fin	Especes fossile	Type fossile	Numéros base de données/article
Groenland	145.5	2.588	<i>Quercus drymeia</i> (A)	Feuilles	(1)
Crête	23.03	5.332	<i>Quercus drymeia</i> (A)	Feuilles	(2)
France/Espagne	40.4	37.2	<i>Quercus drymeia</i> (A)	Non disponible	(3)
Allemagne	28.4	20.4	<i>Quercus drymeia</i> (A)	Feuilles	26765
Grèce	7.246	5.332	<i>Quercus drymeia</i> (A)	Graine	(3)
Italie	7.246	5.332	<i>Quercus drymeia</i> (A)	Non disponible	(3)
Italie	7.246	5.332	<i>Quercus drymeia</i> (A)	Non disponible	(3)
Italie	7.246	5.332	<i>Quercus drymeia</i> (A)	Non disponible	(3)
Italie	7.246	5.332	<i>Quercus drymeia</i> (A)	Non disponible	(3)
Italie	7.246	5.332	<i>Quercus drymeia</i> (A)	Non disponible	(3)
Italie	7.246	5.332	<i>Quercus drymeia</i> (A)	Non disponible	(3)
Italie	7.246	5.332	<i>Quercus drymeia</i> (A)	Non disponible	(3)
Espagne	23.3	20.43	<i>Quercus drymeia</i> (A)	Pollen	(4)
Espagne	5.332	3.6	<i>Quercus drymeia</i> (A)	Non disponible	(3)
République Tchèque	23.03	15.97	<i>Quercus rhenana</i> (B)	Non disponible	(5)
République Tchèque	23.03	15.97	<i>Quercus rhenana</i> (B)	Non disponible	(5)
République Tchèque	23.03	15.97	<i>Quercus rhenana</i> (B)	Non disponible	(5)

République Tchèque	23.03	15.97	<i>Quercus rhenana</i> (B)	Non disponible	(5)
République Tchèque	23.03	15.97	<i>Quercus rhenana</i> (B)	Non disponible	(5)
République Tchèque	23.03	15.97	<i>Quercus rhenana</i> (B)	Non disponible	(5)
Allemagne	16.0	13.7	<i>Quercus rhenana</i> (B)	Feuilles	27520
USA	11.6	1.8	<i>Quercus parvula</i>	Non disponible	23401
USA	11.6	1.8	<i>Quercus remingtonii</i> (C)	Non disponible	23401
USA	1.8	0.0	<i>Quercus brevifolia</i>	Non disponible	34269
<b>USA</b>	<b>46.2</b>	<b>42.0</b>	<b><i>Quercus drymeia</i> (A)</b>	<b>Non disponible</b>	<b>21389</b>
USA	5.3	1.8	<i>Quercus lakevillensis</i> (C)	Feuilles	35272
USA	16.0	11.6	<i>Quercus pseudolyrata</i>	Feuilles	35412
USA	5.3	1.8	<i>Quercus remingtonii</i> (C)	Feuilles	35272
USA	10.3	4.9	<i>Quercus wislizenoides</i>	Feuilles	35271
USA	13.6	10.3	<i>Quercus wislizenoides</i>	Feuilles	27661
USA	13.6	10.3	<i>Quercus wislizenoides</i>	Feuilles	27662
USA	20.4	16.0	<i>Quercus wislizenoides</i>	Feuilles	27179
USA	5.3	1.8	<i>Quercus wislizenoides</i>	Feuilles	35272

(A) : <http://biostor.org/cache/pdf/b9/b1/4c/b9b14c3777da09d710744f77a4097283.pdf>

(B) : Denk T, Grímsson F, Zetter R. Fagaceae from the early Oligocene of Central Europe: Persisting new world and emerging old world biogeographic links. Rev Palaeobot Palyno (2012) 169:7-20.

(C) : <http://books.google.fr/books?id=w095tPDM8WwC&pg=PA7&lpg=PA7&dq=Quercus+lakevillensis&source=bl&ots=nz-7fvv3Ou&sig=HbbogAMhCvs9FS583WEWRZOkItA&hl=fr&sa=X&ei=LImGUK3jBZOZ0FhQff7YDwCg&ved=0CEAQ6AEwBA#v=onepage&q=Quercus%20lakevillensis&f=false>

(1) : <http://www.nrm.se/en/menu/researchandcollections/departments/palaeobotany/collections/databases/schweitzer.9184.html>

(2) : ZIDIANAKIS G., MOHR B. A. R. and FASSOULAS C. 2007. A late Miocene leaf assemblage from Vrysses, western Crete, Greece, and its paleoenvironmental and paleoclimatic interpretation. GEODIVERSITAS, 29 (3), pp. 351-377.

(3) : Kovar-Eeckder, J., Kvacek, Martinetto, ., Roiron, P., 2006, Late Miocene to Early Pliocene vegetation of southern Europe (7-4Ma) as reflected in the megafossil plant record, Palaeogeography, Palaeoclimatology, Palaeoecology 238 (2006) 321-340

(4) : Eduardo Barrón et al, 2010, The Cenozoic vegetation of the Iberian Peninsula: A synthesis, Review of Palaeobotany and Palynology 162 (2010) 382-402

(5) : KVAČEK, Z. & TEODORIDIS, V. 2007. Tertiary macrofloras of the Bohemian Massif: a review with correlations within Boreal and Central Europe. Bulletin of Geosciences 82(4), 383-408 (5 figures, 12 tables). Czech Geological Survey, Prague. ISSN 1214-1119. Manuscript received August 15, 2007, accepted in revised form October 4, 2007, issued December.

L'apport de la présente partie est de trouver pour chacun des six groupes (*Cyclobalanopsis*, *Cerris*, *Ilex*, *Protobalanus*, *Lobatae* et *Quercus s.s*) la date la plus ancienne à travers l'exploration de la base de données FossilMap.

Dans un premier temps, une recherche croisée entre les différentes espèces de fossiles et leur attribution ou non à un des six groupes actuels a été menée. Les espèces fossiles qui n'ont pas d'homologues actuels dans la littérature sont éliminées de la suite de l'analyse. Les espèces fossiles restantes sont donc classées entre elles suivant le groupe actuel auquel elles sont rattachées ; et sont triées en fonction de l'endroit géographique où elles sont répertoriées. Ensuite, la date la plus ancienne est conservée et servira de base pour la calibration des données phylogénétiques. Cependant, lorsque la datation est trop imprécise, par exemple si l'individu est daté du tertiaire (-65 millions d'années à -1.8 ma), cette date ne sera pas prise en compte car l'amplitude est bien trop importante et posera problème pour la suite des analyses.

Les datations des espèces retenues sont toutes basées sur la méthode lithostratigraphique (à chaque strate correspond un âge), à l'exception de la datation utilisée pour *Quercus mediterranea* (Tableau A9) qui provient de données isotopiques et est donc beaucoup plus précise.

De plus, si une espèce est représentée dans les tableaux plusieurs fois avec des dates identiques ou non, cela provient du fait qu'elle s'est retrouvée dans plusieurs localités.

---

#### A4.2.1 DATATION DU GROUPE *CYCLOBALANOPSIS* (TABLEAU A7) :

---

Une seule espèce fossile de *Cyclobalanopsis* a été identifiée (*Cyclobalanopsis naitoi*, Tableau A7) sous forme de feuille. Elle a été trouvée au Japon et est datée de la fin du Paléocène, fin Eocène (-55.8 à -33.9 millions d'années).

Ce groupe est uniquement représenté en Asie, et le fait que l'on ne retrouve des fossiles qu'à cet endroit peut être le témoin de l'évolution locale de ce groupe sans migrations ultérieures.



---

#### A4.2.2 DATATION DU GROUPE *CERRIS* (TABLEAU A8) :

---

Au total, cinq espèces sont attribuées au groupe *Cerris* (Tableau A8) dont une est toujours présente actuellement (*Quercus trojana*). Les espèces fossiles de ce groupe sont uniquement présentes en Europe et en Asie (Tableau A8), tout comme les espèces actuelles. L'espèce la plus ancienne (*Quercus kubinyii*) a été retrouvée en France à l'Eocène (-40.4 à -37.2 millions d'années) et servira à dater l'apparition de ce groupe.

---

#### A4.2.3 DATATION DU GROUPE *ILEX* (TABLEAU A9) :

---

Trois espèces fossiles différentes retrouvées sont attribuées au groupe *Ilex* (Tableau A9) : *Quercus mediterranea*, *Quercus coccifera* et *Quercus ilex* ; les deux dernières espèces étant toujours présentes actuellement. Ces espèces ont été retrouvées uniquement en Europe (Tableau A9), alors que ce groupe est également présent en Asie. Soit les espèces asiatiques n'ont pas été attribuées à ce groupe, soit les archives fossiles sont incomplètes en Asie.

La datation de ce groupe est faite par l'espèce *Quercus mediterranea*, présente en France et en Espagne à l'Eocène (-40.4 à -37.2 millions d'années).

---

#### A4.2.4 DATATION DU GROUPE *QUERCUS S.S* (TABLEAU A10) :

---

Les espèces fossiles associées à ce groupe se retrouvent en Europe, aux Etats-Unis, en Asie (Tableau A10), tout comme leurs membres contemporains.

La plupart de ces espèces sont encore présentes actuellement, ou bien, comme le témoigne leur dénomination (pre + nom de l'espèce) en sont très proches (Tableau A10). L'espèce fossile la plus ancienne est *Quercus prinus* (toujours présente actuellement), retrouvée en Russie et datée de -55.8 à -40.4 millions d'années (Tableau A10). Ceci pourrait suggérer une antériorité de ce groupe en Asie, mais une espèce fossile (*Quercus petros*) a été trouvée aux Etats-Unis datée de -46.2 à -42 millions d'années, soit une période compatible avec l'apparition en Asie.



---

#### A4.2.5 DATATION DU GROUPE *PROTOBALANUS* (TABLEAU A11) :

---

Ce groupe est composé par trois espèces fossiles (*Quercus hannibali*, *Quercus moyei* et *Quercus pliopalmerii*) dont une (*Quercus hannibali*) est très fortement représentée (Tableau A11). Ce groupe, composé actuellement uniquement d'espèces américaines, a des représentants fossiles dans ce continent seulement (Tableau A11). L'espèce servant à la datation de ce groupe est *Quercus moyei*, retrouvée à l'Eocène supérieur (-33.9 à -28.4 millions d'années).

---

#### A4.2.6 DATATION DU GROUPE *LOBATAE* (TABLEAU A12) :

---

Un grand nombre d'espèces fossiles représente ce groupe qui est actuellement présent uniquement aux Etats-Unis, or, bon nombre d'espèces ont été retrouvées en Europe (Tableau A12). Ces observations suggèrent que ce groupe était présent en Europe mais aurait disparu depuis. La théorie selon laquelle les espèces auraient évolué localement sans connaître de disparition qui fonctionnait bien jusque là, peut être mise à mal à travers cet exemple. L'espèce retenue pour dater ce groupe est *Quercus drymeia*, présente à l'Eocène (-46.2 à - 42 millions d'années) aux Etats-Unis (Tableau A12).

---

#### A4.2.7 COMPARAISON ENTRE LES DEUX TYPES DE DATATIONS (TABLEAU A13) :

---

De façon générale, nous remarquons que les datations issues de notre analyse sont différentes de celles de l'équipe de Manos, dans le sens où toutes les datations relevées dans la présente étude y sont antérieures (Tableau A13).

Ceci peut éventuellement poser problème dans les analyses suivantes, bien qu'il ne nous soit pas possible a priori de trancher en faveur de nos datations ou des autres.

Cependant, étant donné que cette étude a prospecté de nombreuses publications et bases de données, il est probable que l'ancienneté des groupes que nous trouvons provienne du fait de notre recherche à été plus exhaustive que les précédentes.

TABLEAU A13: CORRESPONDANCE ENTRE LES DATATIONS RETENUES PAR L'EQUIPE DE MANOS ET LA PRESENTE ETUDE.

<b>Groupe</b>	<b>Datation en millions d'années (Manos)</b>	<b>Datation en millions d'années (Cette étude)</b>
<i>Cyclobalanopsis</i>	ND	33.9-55.8
<i>Cerris</i>	ND	37.2-40.4
<i>Ilex</i>	28.4-33.9	37.2-40.4
<i>Protobalanus</i>	15.3	28.4-33.9
<i>Lobatae</i>	28.4-33.9	42-46
<i>Quercus s.s</i>	28.4-33.9	40.4-55.8

En conclusion, l'analyse de la bibliographie existante sur les fossiles du genre *Quercus* à permis de trouver des datations pour chacun des six groupes (Tableau A13).

La distribution actuelle des espèces incluses dans ces groupes suit bien l'hypothèse selon laquelle les espèces ont évoluées localement, à l'exception du groupe *Ilex* et de façon bien plus étonnante, des *Lobatae* qui se trouvent aujourd'hui uniquement en Amérique du Nord alors que les données fossiles soupçonnent l'existence de ce groupe dans le passé en Europe (Tableau A12).



#### A4.3 REFERENCES BIBLIOGRAPHIQUES

---

- Axelrod, D. I. (1983). "Biogeography of Oaks in the Arcto-Tertiary Province." Annals of the Missouri Botanical Garden **70**(4): 629-657.
- Crepet, W. and C. Daghljan (1980). "Castaneoid inflorescences from the Middle Eocene of Tennessee and the diagnostic value of pollen (at the subfamily level) in Fagaceae." Am J Bot **67**: 739-757.
- Crepet, W. L. (1989). History and implications of the early North American fossil record of Fagaceae. Evolution systematics and fossil history of the Hamamelidae. **2**: 45-66.
- Crepet, W. L. and K. C. Nixon (1989). "Extinct transitional Fagaceae from the Oligocene and their phylogenetic implications." Am J Bot **76**: 1493-1505.
- Daghljan, C. P. and W. L. Crepet (1983). "Oak catkins, leaves and fruits from the Oligocene Catahoula Formation and their evolutionary significance." Am J Bot **70**: 639-649.
- Manos, P. S. and A. M. Stanford (2001). "The Historical Biogeography of Fagaceae: Tracking the Tertiary History of Temperate and Subtropical Forests of the Northern Hemisphere." International Journal of Plant Sciences **162**(S6): S77-S93.
- Meyer, H. and S. Machester (1997). "The Oligocene Bridge Creek flora of the John Day Formation, Oregon." Univ Calif Publ Geol Sci **141**: 1-195.
- Nixon, K. C. and W. L. Crepet (1989). "Trigonobalanus (Fagaceae) - Taxonomic Status and Phylogenetic-Relationships." American Journal of Botany **76**(6): 828-841.
- Zhou, Z.-K. (1992). "Origin, phylogeny, and dispersal of Quercus from China." Acta Bot Yunnanica **14**: 227-236.
- Zhou, Z.-K. (1993). "The fossil history of Quercus." Acta Bot Yunnanica (15): 21-33.



## A5 CONTEXTE, OBJECTIF ET METHODOLOGIE GENERALE DU PRESENT TRAVAIL DE THESE

---

La revue bibliographique sur la phylogénie du genre *Quercus* met en exergue une probable irrésolution des nœuds profonds suggérée par un manque de soutien des branches conduisant à ces nœuds. Tous les marqueurs moléculaires utilisés à ce jour, chloroplastiques ou nucléaires seuls, ou en combinaison, aboutissent à cette conclusion. Cette irrésolution peut venir soit de contraintes techniques, soit de l'histoire évolutive du genre.

Une façon élégante de déterminer si cela est dû à des raisons biologiques, revient à utiliser des marqueurs nucléaires et/ou chloroplastiques indépendants des études précédentes. En effet un événement historique majeur comme une radiation très rapide laissera une empreinte sur l'ensemble du génome, quels que soient les marqueurs moléculaires considérés. L'hypothèse d'une rapide radiation dans le passé, permettrait de comprendre l'absence de signal phylogénétique observée chez plusieurs gènes. Dans ce cas, on devrait explorer un nombre important de marqueurs mutants lentement, en espérant que l'ensemble de l'information portée par ces marqueurs produise un signal majoritaire en adéquation avec ce phénomène.

Ceci sera testé grâce à un échantillonnage de 110 espèces sur le génome nucléaire à travers deux types de marqueurs : les marqueurs conservés entre espèces ou COS (partie B2 du manuscrit), et des gènes nucléaires issus d'un projet de reséquençage chez *Quercus robur*, *Q. petraea* et *Q. Ilex* (partie B3 du manuscrit). Le génome chloroplastique total (>160 000 pb) sera étudié sur un sous-échantillon de 57 espèces (partie B4).

La comparaison de ces différents types de marqueurs et de génomes devrait nous conforter ou infirmer notre hypothèse de radiation rapide : dans ce cas, chaque gène ne sera pas porteur d'information au niveau des nœuds profonds, mais l'ensemble des gènes le sera. D'autre part la comparaison des phylogénies entre génome nucléaire et cytoplasmique devrait également nous aider à comprendre le rôle de leur hérédité particulière dans le signal phylogénétique observé (partie B5). L'utilisation de la base de données fossile FossilMap permettra de calibrer (dater) les nœuds sur les phylogénies nucléaires et chloroplastiques obtenues, et nous aidera à comprendre la biogéographie historique du genre *Quercus*.



---

PARTIE B

RECHERCHE DE MARQUEURS

MOLECULAIRES

---



FIGURE B1A : LOCALISATION DES ESPECES ECHANTILLONNEES SUR LE CONTINENT AMERICAIN.



FIGURE B1B : LOCALISATION DES ESPECES ECHANTILLONNEES SUR LE CONTINENT EURASIATIQUE.



## B1 MATERIEL ET METHODES MOLECULAIRES :

---

Le matériel biologique ainsi que les méthodes de reconstructions phylogénétiques sont communes pour les deux types de marqueurs nucléaires étudiés. Les Matériel et Méthodes seront donc évoqués de manière générique dans cette partie, à laquelle il sera fait référence par la suite. Concernant le séquençage du génome chloroplastique, l'échantillonnage sera décrit dans la partie le concernant (partie B4), les méthodes également bien que la méthodologie utilisée soit de même nature que celle présentée ci-dessous.

### B1.1 ECHANTILLONNAGE

---

La présente étude fait suite en quelque sorte à une thèse soutenue en 2004 (Xu, 2004) et une partie des échantillons avait été récoltée à l'époque. Différentes récoltes indépendantes ont été menées lors de missions au Mexique, ou faites par des collègues généticiens forestiers en Thaïlande, en Israël, en Iran, en Turquie et au Costa Rica. Ces récoltes complémentaires sont venues consolider l'échantillon initialement constitué pour la thèse de Xu Li An. Enfin, un dernier complément a été réalisé au début de ma thèse par des récoltes faites dans l'arboretum des Pouyouleix en Dordogne ([http://www.jardinez.com/Parcs-et-jardins-Arboretum-des-Pouyouleix\\_Saint-Jory-de-Chalais\\_Dordogne\\_Aquitaine\\_fr\\_14974.html](http://www.jardinez.com/Parcs-et-jardins-Arboretum-des-Pouyouleix_Saint-Jory-de-Chalais_Dordogne_Aquitaine_fr_14974.html)). Cet Arboretum est de constitution récente et les origines des arbres qui y sont installés sont bien documentées.

L'échantillonnage sur le terrain s'est fait toujours de la même façon : des feuilles (ou bourgeons) ont été récoltées sur un à sept arbres par espèce séparés de moins de 500 mètres (afin de limiter le polymorphisme entre espèces éloignées géographiquement) et mises immédiatement en présence de silicagel afin de les sécher, puis elles ont été conservées à -80°C en vue de l'extraction de leur ADN.

Au total, 111 espèces de *Quercus*, de *Notholithocarpus* et de *Castanea* forment l'échantillon utilisé lors des travaux de cette étude (Figure B1A et B1B, Tableau B1 et Annexe 1).

Si on s'intéresse aux *Cyclobalanopsis*, neuf espèces ont été échantillonnées sur un total compris entre 76 et 87 en fonction des auteurs (Tableau B1), mais ces espèces balayent bien la variabilité taxonomique existante.



Les *Quercus s.s* sont très bien représentés car constitués de six espèces asiatiques sur un total de 15-25 (Tableau B1), huit espèces en Europe et Afrique du Nord sur un total de 16 (Tableau B1), et 35 espèces Américaines sur un total de 104 (Tableau B1). Plus de un tiers des espèces de *Lobatae* a été échantillonné (Tableau B1), quatre sur cinq espèces de *Protobalanus* ont été récoltées et 10 espèces du groupe *Cerris* sur un total de 12 (Tableau B1). Les deux espèces du groupe *Ilex* d'Europe et d'Afrique du Nord ont été récoltées (Tableau B1), mais seulement trois sur les 26 espèces Asiatiques l'ont été, ceci peut malheureusement constituer un trou dans notre échantillonnage.

---

## B1.2 EXTRACTION D'ADN, AMPLIFICATION ET SEQUENÇAGE

---

L'ADN génomique total a été extrait en utilisant le kit commercial Invisorb DNA plant HTS en plaque 96 en utilisant les recommandations du fabricant.

La qualité de l'ADN a été vérifiée visuellement par gel d'agarose 1% (poids/volume) coloré au GelRed (Biotium, USA). La concentration a ensuite été ajustée à 10ng/μl après quantification sur un spectromètre à 8 canaux. Les réactions de PCR se sont faites dans des volumes de 20 μl contenant 2 μl de tampon 10X, 5mM de MgCL<sub>2</sub>, 5mM de dNTP, 4μM de chaque amorce et 0.5 U d'Amplitaq Gold 360 (Applied biosystem) en présence de 20 ng d'ADN.

TABLEAU B1 : NOMBRE D'ESPECES COLLECTEES ET ANALYSEES ET NOMBRE TOTAL D'ESPECES DU GENRE QUERCUS REPERTORIEES DANS LA LITTERATURE.

Genre	Sous-genre	Groupe	Origine Géographique	Nombre d'espèces échantillonnées	Nombre d'espèces estimée
<i>Castanea</i>			Asie	1	12
<i>Notholithocarpus</i>			Amérique	1	1
<i>Quercus</i>					
	<i>Cyclobalanopsis</i>		Asie	9	87e, 76c
	<i>Quercus s.l.</i>				
		<i>Quercus s.s.</i>	Asie	6	25e, 15c
			Europe et Afrique du Nord	8	16d
			Amérique	35	104a
		<i>Lobatae</i>	Amérique	31	82a
		<i>Protobalanus</i>	Amérique	4	5b
		<i>Cerris</i>	Asie	6	8c
			Europe et Afrique du Nord	4	4d
		<i>Ilex</i>	Asie	3	26c
			Europe et Afrique du Nord	2	2d
				109	

(a) Nixon K.C., 1993b.

(b) Floras (Brach and Song, 2006),( <http://www.efloras.org> ).

(c) Menitsky Y.L. 2005.

(d) Flora europea (<http://rbg-web2.rbge.org.uk/FE/fe.html>).

(e) Camus (1936-1954).

Le même programme de PCR a été appliqué sur tous les marqueurs de la façon suivante :

- Activation de l'enzyme à 95°C pendant 10 minutes,
  - Dénaturation à 94°C pendant 1 minute
  - Hybridation à 59°C pendant 1 minute
  - Elongation à 70°C pendant 2 minutes
- } X2
- 
- Dénaturation à 93°C pendant 45 secondes
  - Hybridation à 59°C pendant 30 secondes (-0.5° par cycle)
  - Elongation à 70°C pendant 2 minutes
- } X17
- 
- Dénaturation à 92°C pendant 30 secondes
  - Hybridation à 50°C pendant 30 secondes
  - Elongation à 70°C pendant 2 minutes
- } X19
- 
- Elongation finale à 70°C pendant 10 minutes

Tous les fragments ont été séquencés sur la plateforme Génome-Transcriptome de Bordeaux sur un séquenceur automatique à capillaires ABI 3730. Les séquences ont été éditées via le logiciel CodonCode Aligner version 3.7 (CodonCode Corporation, Dedham, MA, USA), les séquences n'amplifiant pas en forward et en reverse ont été éliminées. Les séquences ont ensuite été alignées avec l'algorithme Muscle de la version 4 du logiciel Seaview (Gouy *et al.*, 2010). Lorsque plusieurs individus d'une même espèce ont donné une séquence, une séquence de consensus majoritaire a été construite avec le logiciel Bioedit v7.1.3 (Hall, 1999).



### B1.3 RECONSTRUCTIONS PHYLOGENETIQUES

---

Afin d'aider au maximum le lecteur, et de lui permettre d'appréhender plus facilement la suite du manuscrit, je me permets de revenir sur certains points couramment énoncés en phylogénie et qui guident la méthodologie employée dans chacune des parties suivantes.

Une phrase qui revient souvent en phylogénie moléculaire est : « L'arbre des gènes n'est pas l'arbre des espèces ». En effet, une espèce ne se restreint pas à ses gènes, car ceux-ci ont une histoire propre. Comparer différents gènes entre eux peut donner lieu à des contradictions, des incongruences car leur histoire peut être différente.

Si tous les arbres issus des marqueurs trouvés sont porteurs de signal phylogénétique et sont congruents entre eux, dans ce cas, il est possible de les interpréter et de les utiliser ensemble (en utilisant l'information topologique portée par chacun des arbres à l'aide de la méthode dite de SuperArbre; ou bien en assemblant les données entre elles sous la forme d'une SuperMatrice qui sera intégrée dans un logiciel de phylogénie).

L'approche SuperArbre consiste à obtenir un arbre phylogénétique pour chaque gène, puis de faire une analyse globale suivant certains critères, de manière à générer un SuperArbre qui tient compte des données topologiques fournies par chaque arbre individuel.

L'approche alternative dite de SuperMatrice consiste à assembler tous les gènes en une SuperMatrice qui servira de base ensuite à la production d'un seul arbre phylogénétique. Cet assemblage se fait en concaténant toutes les séquences. Il est alors possible que le signal phylogénétique soit minoritaire sur chaque gène mais que lorsque l'on les concatène, ce signal devienne majoritaire et que donc le résultat final soit un arbre plus résolu et plus soutenu que lors de l'utilisation de la méthode de SuperArbre.

Ces deux types de représentations présentent l'avantage de pouvoir intégrer des données manquantes (= des espèces manquantes) dans l'arbre final.

Dans le cas contraire où les arbres de gènes sont incongruents (des clades soutenus par des valeurs de bootstraps élevés sont incompatibles entre les différents arbres), trois phénomènes sont classiquement évoqués :



\* Il est probable que les gènes étudiés soient des gènes paralogues issus de duplication du génome puis de la perte d'une des copies (comme ce fut le cas pour les ITS provenant d'études concurrentes);

\* Il se peut qu'il y'ait eu des transferts horizontaux (hybridations) et que dans ce cas, il ne soit pas possible de comparer des arbres dichotomiques entre eux;

\* Il peut y avoir eu tri incomplet des lignées ancestrales (également appelé coalescence profonde). Le polymorphisme est principalement hérité après spéciation par les deux espèces filles, mais il est possible que les différentes espèces fixent le polymorphisme de manière indépendante et donc que l'histoire des allèles ne soit pas la même que l'histoire des organismes.

En résumé, pour chacun des trois marqueurs moléculaires étudiés (deux types différents de marqueurs nucléaires en plus des marqueurs chloroplastiques) la même méthodologie sera appliquée pour chacun des jeux de données :

\* Il faudra s'assurer que chaque gène individuel soit compatible avec les autres (pas d'incongruence topologique),

\* Si tel est le cas, la méthode de SuperArbre et de SuperMatrice sera appliquée sur l'ensemble des gènes. Si la phylogénie obtenue est satisfaisante, ces marqueurs seront gardés pour la suite, dans le cas contraire, ils seront éliminés des analyses suivantes.

\* Chaque jeu de gènes sera comparé aux autres, s'ils sont compatibles entre eux, ils seront analysés conjointement (par une méthode de SuperArbre et/ou de SuperMatrice) ou si ce n'est pas le cas, les hypothèses comme l'hybridation seront évoquées.

Les phylogénies gène par gène ont été obtenues par Maximum de Vraisemblance par le logiciel PhyML v.3.0 (Guindon and Gascuel, 2003) en appliquant les modèles d'évolution obtenus (Tableau B4) par le logiciel ModelTest 3.7 (Posada and Crandall, 1998). Les valeurs de supports des branches (bootstraps) ont été obtenues après 100 ré échantillonnage.

Pour l'analyse SuperMatrice (concaténation de tous les marqueurs), la concaténation de tous les gènes a été réalisée avec le logiciel Seaview (version 4).

Dans le cas où les marqueurs définis sont gardés dans la suite des analyses, deux types de reconstructions phylogénétiques seront menées : une par Maximum de Vraisemblance (MV)



avec le logiciel RAxML 7.2.8-ALPHA (Stamatakis, 2006), et une par Inférence Bayésienne (BI) avec le logiciel MrBayes 3.1.2 (Ronquist and Huelsenbeck, 2003) ; sinon, seule la recherche par Maximum de Vraisemblance est effectuée.

Concernant la recherche par Maximum de Vraisemblance, celle-ci suit les recommandations de l'auteur, à savoir que la recherche se fait en optimisant le modèle d'évolution GTR + Gamma sur chaque gène de façon indépendante, avec 4 types différents de vitesse d'évolution des sites. 1000 ré échantillonnages par bootstraps sont menés en parallèle.

L'Inférence Bayésienne a été menée uniquement si les gènes ont été déclarés comme phylogénétiquement informatifs et consiste en deux runs parallèles de 20 millions de générations avec échantillonnage tous les 1000 pas. Chaque run étant constitué de 2 types de chaînes : 3 chaînes chaudes et une chaîne froide. La convergence est atteinte lorsque la déviation standard moyenne entre les runs est inférieure à 0.05.

Dans le but de simplifier au maximum la lecture des nombreux arbres que nous allons voir, le code couleur suivant sera appliqué : le marron correspondra aux groupes utilisés pour raciner les arbres, le jaune aux *Cyclobalanopsis*, le rose au groupe des *Cerris*, le vert au groupe *Ilex*, le rouge aux *Lobatae*, le bleu aux *Protobalanus* et le gris aux *Quercus s.s.*



## B2 CONSERVED ORTHOLOG SET MARKERS (COS)

---

### B2.1 OBJECTIF

---

Comme nous l'avons vu dans la partie précédente, le but de la recherche de nouveaux gènes est de voir si une meilleure résolution de la phylogénie peut être obtenue au niveau des nœuds profonds avec des valeurs de soutien de branches élevés.

Nous avons profité du développement au sein du laboratoire de deux types de marqueurs nucléaires : un jeu de gènes COS (Conserved Ortholog Set markers) et un jeu de gènes de ménage qui leur feront suite.

Les COS sont des gènes en copies uniques ou présents en peu de copies, orthologues et conservés entre espèces. On parle d'orthologie entre un gène chez deux espèces, quand ce gène est transmis dans chaque espèce après spéciation. Ce terme est à mettre en opposition avec le terme de paralogie. Lorsqu'un gène est dupliqué au sein d'un génome, les deux copies peuvent évoluer de façon indépendante, après spéciation, une espèce peut hériter d'une des copies et l'autre de l'autre copie. Les deux gènes, qualifiés de paralogues dans ce cas, pourront avoir une histoire différente et ne devront pas être utilisés en phylogénie !

Nous avons vu un cas de paralogie dans la partie introductive lorsque nous avons cité les travaux de Samuel sur les ITS (Samuel *et al.*, 1998).

En l'absence de données sur le séquençage complet du génome, une façon de voir si des gènes sont paralogues ou orthologues est de voir le degré de similitude entre espèces du gène en question, voir s'il y a des codon stop au sein de la zone transcrite (suggérant une copie non fonctionnelle du gène) ou leur position sur une carte génétique.

Un même gène très polymorphe au niveau de la séquence nucléotidique chez deux espèces suggérera une paralogie, alors que nous serons confiants face à deux gènes quasi identiques chez deux espèces.



Cependant, un phénomène résurgent peut être évoqué : la paralogie cachée. En effet, si la spéciation a lieu peu de temps après la duplication par exemple, cette technique de détection s'avèrera inefficace.

Les COS sont donc des gènes que l'on suppose orthologues entre espèces mais également conservés entre espèces lointaines. En effet, dans une phylogénie multilocus, il s'avère important d'avoir des gènes qui mutent lentement et d'autres plus rapidement dans le but de balayer le maximum d'événements évolutifs (les spéciations anciennes et plus récentes). Avoir des gènes conservés dans des espèces lointaines (de la vigne *Vitis vinifera* au chêne *Quercus robur*) permet en théorie d'avoir des gènes qui ont peu muté et sont donc informatifs des événements de spéciation anciens, c'est dans ce cadre là que les COS ont été choisis.

---

## B2.2 CHOIX DES MARQUEURS

---

Le travail bioinformatique de recherche de gènes COS a été réalisé par Yao-Cheng Lin au « VIB Department of Plant Systems Biology » à l'Université de Ghent en se basant sur la recherche de gènes orthologues (Ueno *et al.*, 2010).

La recherche bio-informatique des COS s'est faite sur deux jeux de données : six génomes de plantes complets (*Arabidopsis thaliana* et *lyrata*, *Populus trichocarpa*, *Medicago truncatulata*, *Glycine max* et *Vitis vinifera*) et trois génomes incomplètement séquencés dont seules des données EST (Expressed Sequence Tag) étaient disponibles : *Quercus*, *Castanea* et *Fagus*.

Une représentation phylogénétique simplifiée ainsi que les différentes catégories de COS qui ont été faites sont indiquées sur la figure B2. Les lettres allant de A à N indiquent les différentes catégories de COS qui ont été produits, à chaque ligne correspond une espèce qui est placée sur l'arbre simplifié à gauche, chaque valeur correspond au nombre toléré de copie du gène (V voulant dire qu'il n'y a aucune restriction sur le nombre de copies de gènes, et 1-5 exprimant la possibilité qu'un EST puisse se retrouver affecté plusieurs fois sur un même gène).

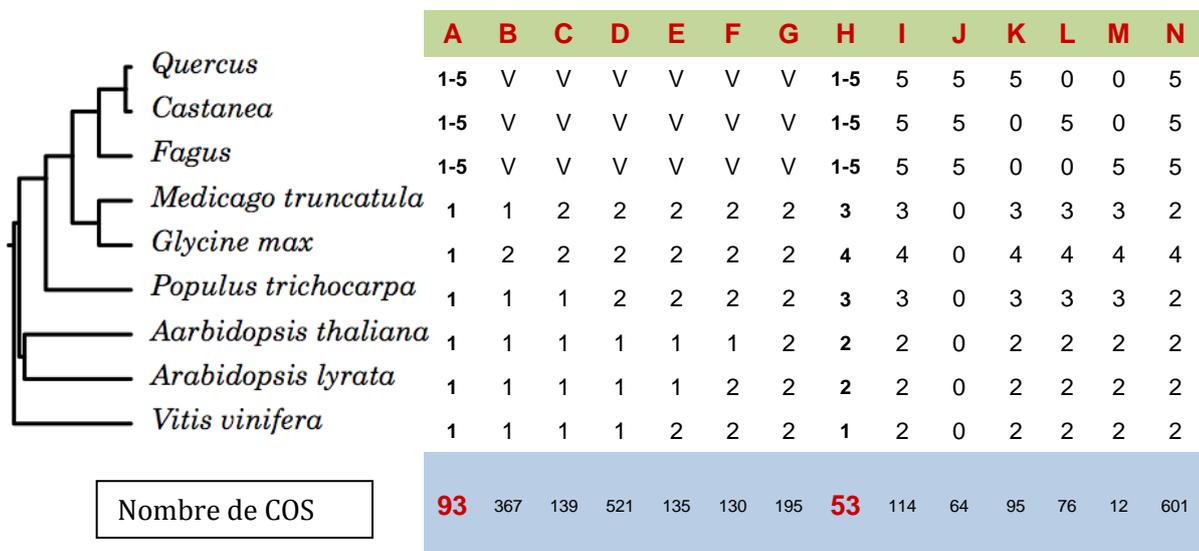


FIGURE B2 : REPRESENTATION PHYLOGENETIQUE SIMPLIFIEE DES 9 GENOMES ETUDIES DANS LE CADRE DE LA RECHERCHE DE COS. CHAQUE LETTRE ALLANT DE A A N CORRESPOND A UNE CATEGORIE DE COS. LES CHIFFRES CORRESPONDENT AU NOMBRE DE COPIES DE CHAQUE COS DANS LE GENOME ETUDIE QUI EST MIS EN CORRESPONDANCE DES COLONNES DANS L'ARBRE PHYLOGENETIQUE, V INDIQUANT QU'AUCUNE RESTRICTION N'EST EMISE.

Dans le cadre de notre étude, nous avons décidé de concentrer nos efforts sur des COS dont une seule copie était présente chez tous les génomes entièrement séquencés (Colonne A, Figure B2), avec un nombre d'EST compris entre un et cinq (et pouvant se retrouver affecté plusieurs fois sur un même gène à cause du phénomène d'épissage alternatif chez les espèces non entièrement séquencées); mais également en éliminant la contrainte du nombre de copies chez les espèces entièrement séquencées (Colonne H, Figure B2).

Ainsi, 93 COS correspondant aux contraintes émises en A et 53 émises en H ont été identifiés, soit un total de 146 gènes potentiellement utilisables en phylogénie.

Plusieurs filtres ont été appliqués sur ce jeu de gènes. En effet, 19 séquences s'alignaient mal entre elles et ont été éliminées, neuf ne donnaient aucun résultat de blast sur la base de données du peuplier ([http://genome.jgi-psf.org/cgi-bin/runAlignment?db=Poptr1\\_1&advanced=1](http://genome.jgi-psf.org/cgi-bin/runAlignment?db=Poptr1_1&advanced=1)), 64 étaient présentes en multiples copies sur la base de données du peuplier (les paramètres par défauts étaient sélectionnés) et, au final, 54 COS ont été utilisés dans le but de les amplifier par PCR. Sur ces 54 COS, des amorces ont pu être dessinées sur la moitié d'entre eux. Ensuite, pour qu'un marqueur soit valide, il fallait qu'il amplifie sur au moins un des deux individus par groupe d'un panel ; ce panel était constitué de deux espèces de chaque groupe *Castanea*, *Cyclobalanopsis*, *Quercus s.s*, *Lobatae*, *Cerris*, *Ilex* et *Protobalanus*. En effet, étant donné que les reconstructions phylogénétiques se focalisent sur les six groupes définis préalablement, il nous est apparu nécessaire que l'amplification par PCR puisse se faire aisément sur tous les groupes afin d'éviter d'avoir des groupes manquants.

Sur ces 27 COS restants, 14 ont donné des bandes multiples sur gel d'agarose et ont été éliminés, neuf séquences n'étaient pas utilisables car les séquences étaient de mauvaise qualité. Au total, quatre COS ont été définis dont les propriétés sont données dans le tableau B2 et les séquences dans le tableau B3.

La liste des individus par gène qui ont amplifié est donnée dans l'annexe 1.

TABLEAU B2 : CARACTERISTIQUES DES GENES SELECTIONNES SUR L'ENSEMBLE DES ESPECES DE *QUERCUS* ECHANTILLONNES ET DE *CASTANEA MOLLISSIMA*.

	<b>Annotation</b>	<b>Partie transcrite (pb)</b>	<b>Partie non transcrite (pb)</b>	<b>Nombre de sites polymorphes / taille</b>	<b>Nombre de sites parcimonieux / taille</b>
<b>A5799</b>	Molybdopterin synthase sulfur carrier subunit	318	407	0.23	0.15
<b>A6785</b>	Protein slowmo homolog	389	-	0.19	0.15
<b>H10808</b>	No hit	316	-	0.29	0.198
<b>H9687</b>	Zinc finger CCCH domain-containing protein 65	96	381	0.31	0.23

TABLEAU B3 : SEQUENCES DES AMORCES UTILISEES POUR L'AMPLIFICATION DES MARQUEURS COS

	<b>Séquence de l'amorce forward</b>	<b>Séquence de l'amorce reverse</b>
<b>A5799</b>	TAGGGCCACTATGGGAAATG	CAAAAGCTATAAGTCGTAAGCAGGA
<b>A6785</b>	CATAATGGTTAGAGCATACACTCAAG	GGGACCCTTCTTGTTTACAC
<b>H10808</b>	AAGCTGTGTGTTAAGGCAGATG	GGAGAATTTGCAATCAAAGTATC
<b>H9687</b>	GCTGTCATTTTGCACGTCAC	GGAAGCAGGCAAGACATC

TABLEAU B4 : MODELES D'EVOLUTION DES GENES RETENU POUR CHAQUE MARQUEUR COS

	<b>COS A6785</b>	<b>COS A5799</b>	<b>COS H9687</b>	<b>COS H10808</b>
<b>Modèle sélectionné</b>	HKY	HKY + G	HKY + G	HKY+I

HKY : Modèle d'évolution à cinq paramètres de Hasegawa, Kishino et Yano.

G : Optimisation des taux d'évolutions par la loi Gamma.

I: Optimisation des taux d'évolutions en incluant la possibilité d'avoir des sites invariants.



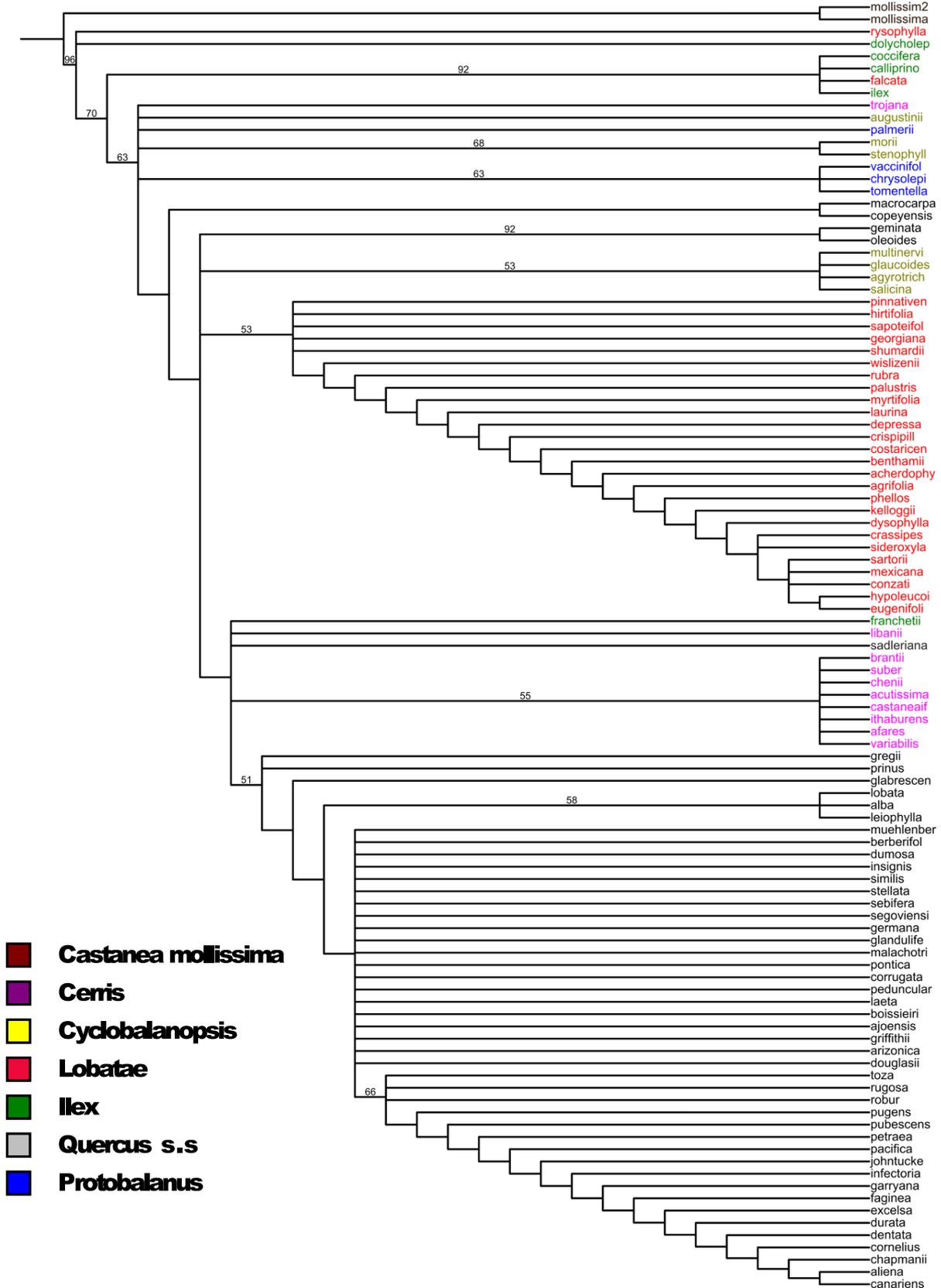


FIGURE B3: REPRESENTATION PAR MAXIMUM DE VRAISEMBLANCE DU COS 6785 SANS LONGUEUR DE BRANCHE, SEULS LES BOOTSTRAPS SUPERIEURS A 50% SONT INDIQUES.

## B2.3 RESULTATS GENE PAR GENE

---

Les différents arbres ont été racinés sur le genre *Castanea*, seules les valeurs supérieures à 50% apparaissent, les nœuds de valeurs inférieurs sont représentés sous forme de polytomies, les longueurs de branches ne sont pas prises en compte et les arbres ont été visualisés à l'aide du logiciel en ligne Interactive Tree Of Life version 2.0 (<http://itol.embl.de>).

---

### DESCRIPTION DU COS A6785 (FIGURE B3)

---

La figure B3 représente l'arbre phylogénétique obtenu à partir de toutes les séquences qui ont donné un résultat sur le COS A6785, seules les valeurs de bootstraps supérieures à 50% sont indiquées afin de simplifier la vue d'ensemble. Ce gène qui a une longueur de 389 pb (Tableau B2) est associé à la fonction de « Protein slowmo homolog », il code pour une protéine mitochondriale de fonction inconnue chez la Drosophile.

Il y a peu de branches bien soutenues ; les grands groupes, bien que peu soutenus, semblent respectés à l'exception du groupe des *Ilex* (*Quercus ilex*, *Quercus coccifera*, *Quercus calliprinos*) qui contient un *Lobatae* (*Quercus falcata*) avec un bootstrap fort (BP 92%).

Ce résultat semble assez étrange, en effet, il n'y a pas de structure phylogénétique nette dans cet arbre. Au-delà de raisons biologiques inconnues, nous pourrions penser que *Quercus falcata* serait une contamination ou que bien le phénomène de tri incomplet des lignées ancestrales (ILS) ait conclu à cette incongruence.

---

### DESCRIPTION DU COS A5799 (FIGURE B4)

---

Ce gène de 725 pb est constitué d'une partie transcrite de 318 pb et d'une partie non transcrite de 407 pb (Tableau B2). Selon son annotation, il code pour une Molybdopterin synthase sulfur carrier subunit. Ce gène a pour fonction de transporter le soufre nécessaire à la synthèse de la molybdoptérine chez *Arabidopsis thaliana* (Mayer *et al.*, 1999).

La racine, les *Protobalanus*, les groupes *Cerris* et *Ilex*, les *Lobatae* et les *Quercus s.s* sont très bien soutenus (bootstraps supérieurs à 80% sauf pour le groupe *Cerris* où la valeur est de 64%).

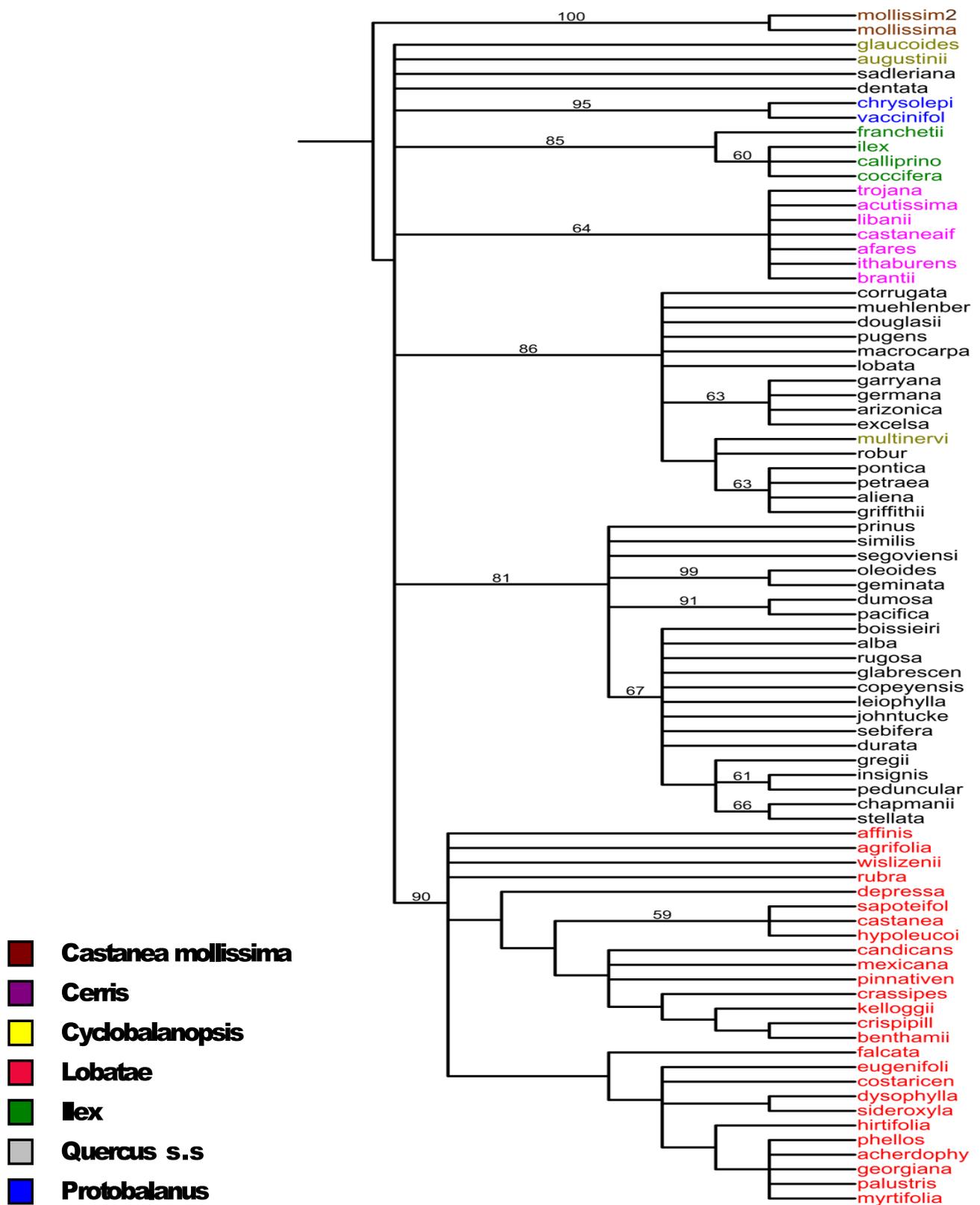


FIGURE B4 : REPRESENTATION PAR MAXIMUM DE VRAISEMBLANCE DU COS 5799 SANS LONGUEUR DE BRANCHE, SEULS LES BOOTSTRAPS SUPERIEURS A 50% SONT INDIQUES.

Les *Quercus s.s* sont séparés en deux groupes bien soutenus, bien qu'aucune relation biologique ne puisse être dégagée. Il existe cependant une espèce de *Cyclobalanopsis (Quercus multinervis)* qui dans ce cas se retrouve dans un groupe inadéquat.

Comme précédemment, aucune résolution n'apparaît au niveau des nœuds profonds : cet arbre phylogénétique est caractérisé par une polytomie.

---

#### DESCRIPTION DU COS H10808 (FIGURE B5)

---

Ce COS d'une longueur de 318 pb est uniquement composé d'une partie transcrite (Tableau B2). Aucune annotation n'a été trouvée. Il est caractérisé par très peu de groupes soutenus et aucune relation phylogénétique intergroupe n'apparaît.

---

#### DESCRIPTION DU COS H9687 (FIGURE B6)

---

La partie transcrite de ce gène fait une taille de 96pb, la partie non transcrite 381 pb. Ce gène est annoté comme étant une Zinc finger CCCH domain-containing protein 65. Le rôle de cette protéine est d'hydrolyser les ponts phosphodiester des acides nucléiques chez *Arabidopsis thaliana* (Sato *et al.*, 2000). Les *Quercus s.s* sont bien soutenus (bootstrap de 74%), bien que deux espèces se retrouvent à part et forment un groupe soutenu (*Quercus pontica* et *Quercus sadleriana*) avec une valeur de 88%. Le regroupement de ces deux espèces était également très fortement soutenu dans le COSH10808 (Figure B5).

Il est intéressant de noter à ce stade que d'autres auteurs ont également retrouvés ces deux espèces à part (Denk and Grimm, 2010) et ont conclu que ceci provenait du fait que ces espèces sont certainement des espèces reliques du tertiaire (-65 à -1.8 millions d'années).

Les *Lobatae* se retrouvent scindés en trois groupes soutenus, avec quelques espèces isolées.

Il n'y a pas de structure phylogénétique apparente dans ce gène.

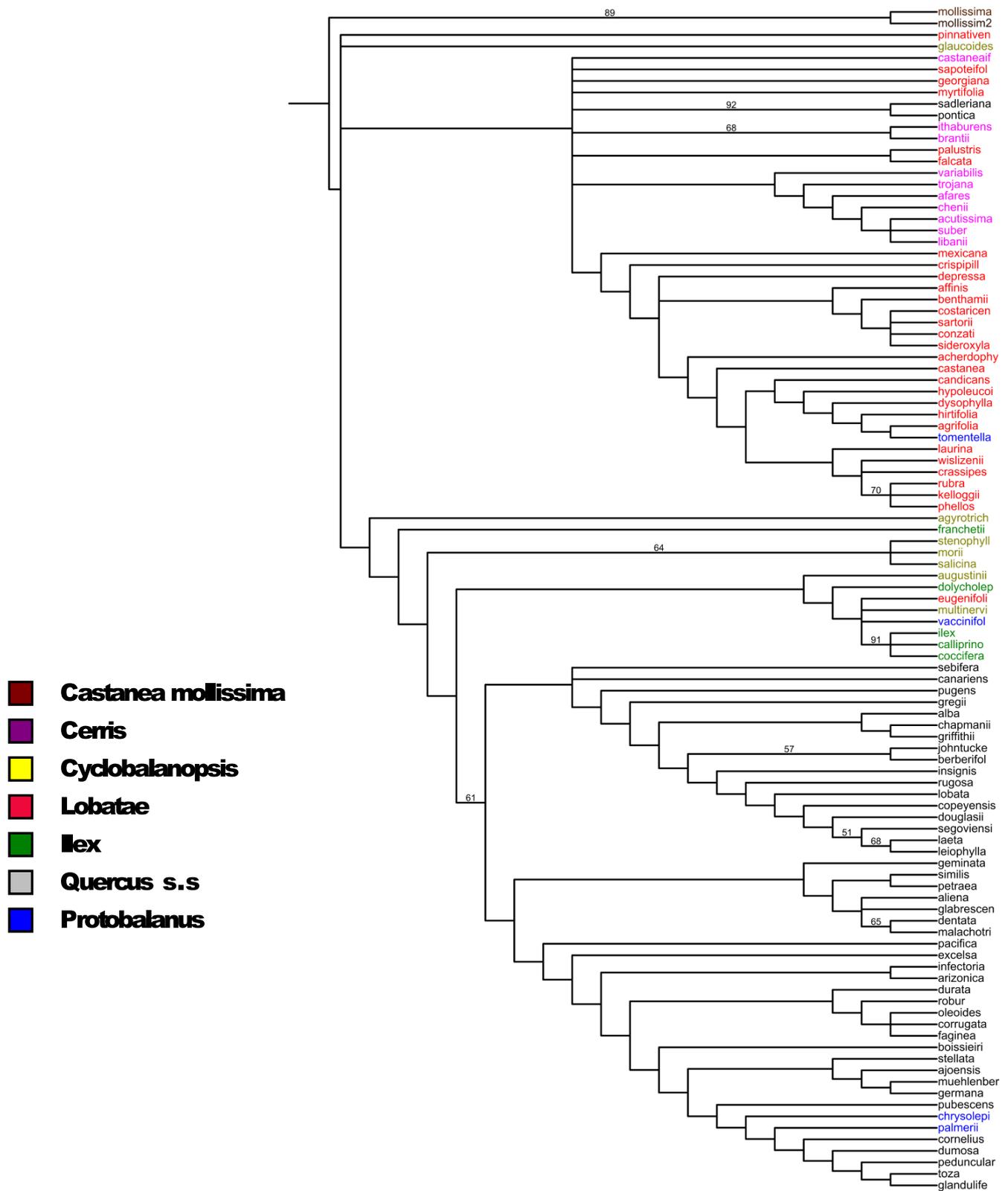


FIGURE B5 : REPRESENTATION PAR MAXIMUM DE VRAISEMBLANCE DU COS 10808 SANS LONGUEUR DE BRANCHE, SEULS LES BOOTSTRAPS SUPERIEURS A 50% SONT INDICES.

## B2.4 RESULTAT SUPER ARBRE (FIGURE B7)

---

Lorsque dans plusieurs arbres phylogénétiques individuels les mêmes taxons sont représentés, la façon la plus simple de représenter un SuperArbre est de faire un consensus qui peut être de type strict ou majoritaire des différents taxa constituant les clades.

Il existe bon nombre d'algorithmes construisant des SuperArbres mais le détail de ces méthodes n'est pas du ressort de la présente thèse, nous avons choisi de présenter les résultats d'une méthode basée sur les triplets : la méthode Supertriplets.

L'arbre issu de cette méthode (Figure B7) n'est pas du tout satisfaisant. En effet, il ne fait apparaître aucune relation phylogénétique entre les clades. De plus, de grosses incohérences subsistent et sont soutenues assez fortement comme la présence d'un *Cyclobalanopsis* (*Quercus salicina*) au sein des *Quercus s.s.*

A ce stade, nous pouvons conclure par le fait que la méthode SuperArbre échoue à reconstruire une phylogénie correcte à partir des arbres sources qui ont été donnés, sans doute par manque de signal phylogénétique disponible dans chacun des quatre gènes.

## B2.5 RESULTAT SUPER MATRICE (FIGURE B8)

---

Le modèle sélectionné par Modeltest est HKY + G + I et a été inclus dans les paramètres d'entrées du logiciel de reconstruction phylogénétique PHYML v3. après concaténation des quatre COS en une SuperMatrice. La figure B8 représente cet arbre.

Sur les 110 taxa, cinq seulement ne se retrouvent pas affectés à un groupe particulier; tous les autres taxa le sont avec des valeurs de bootstraps plus ou moins élevées allant de 58% à 100%.

Aucun taxa n'est dans un groupe inadéquat vis-à-vis des groupes botaniques connus, contrairement à ce qui a été observé dans le SuperArbre précédent (Figure B7).

Ces résultats encourageants sont cependant à nuancer. En effet, bien que tous les groupes soient reconnus et assez bien soutenus, aucune relation phylogénétique entre ces groupes n'apparaît dans cette analyse.

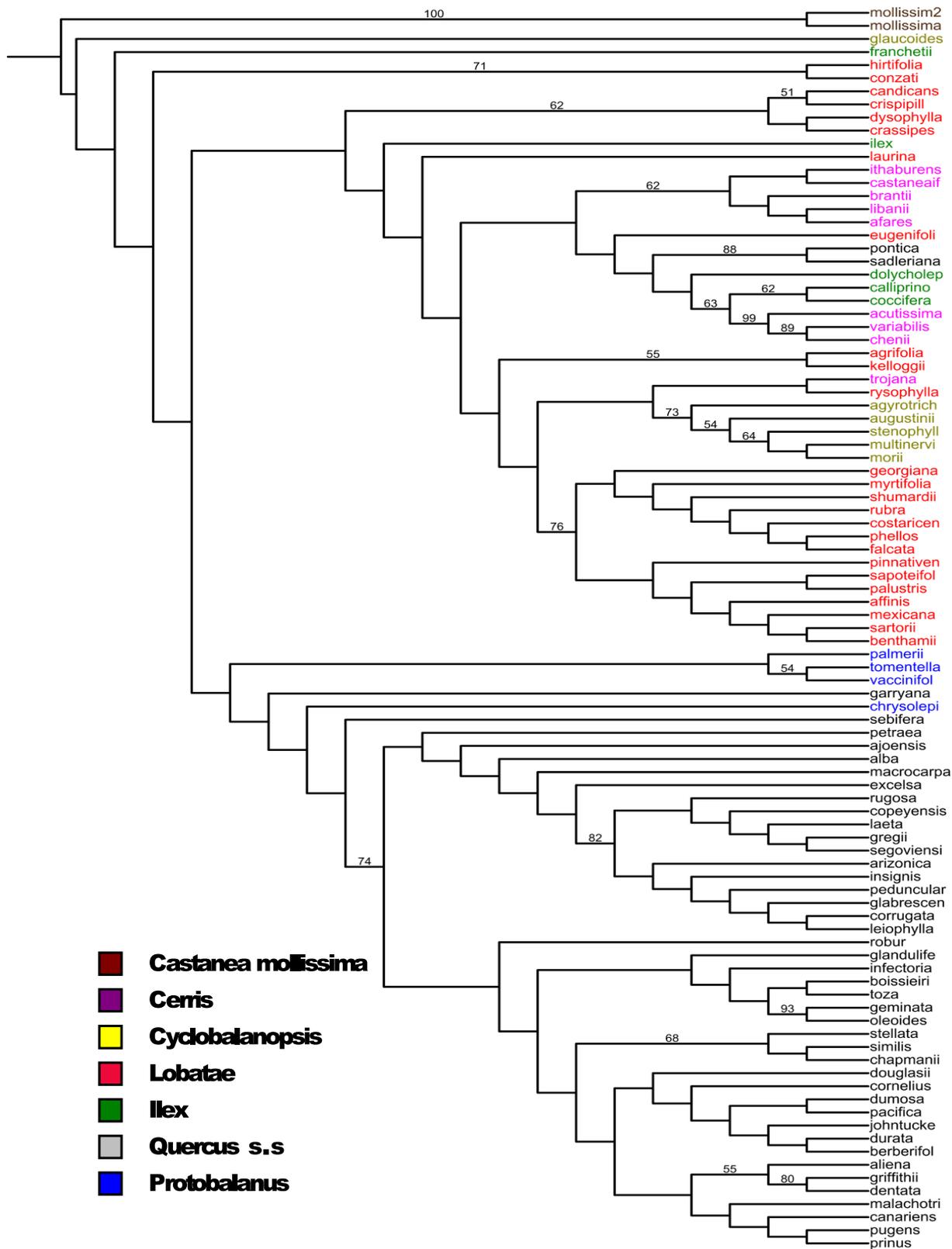


FIGURE B6 : REPRESENTATION PAR MAXIMUM DE VRAISEMBLANCE DU COS 9687SANS LONGUEUR DE BRANCHE, SEULS LES BOOTSTRAPS SUPERIEURS A 50% SONT INDICUES.



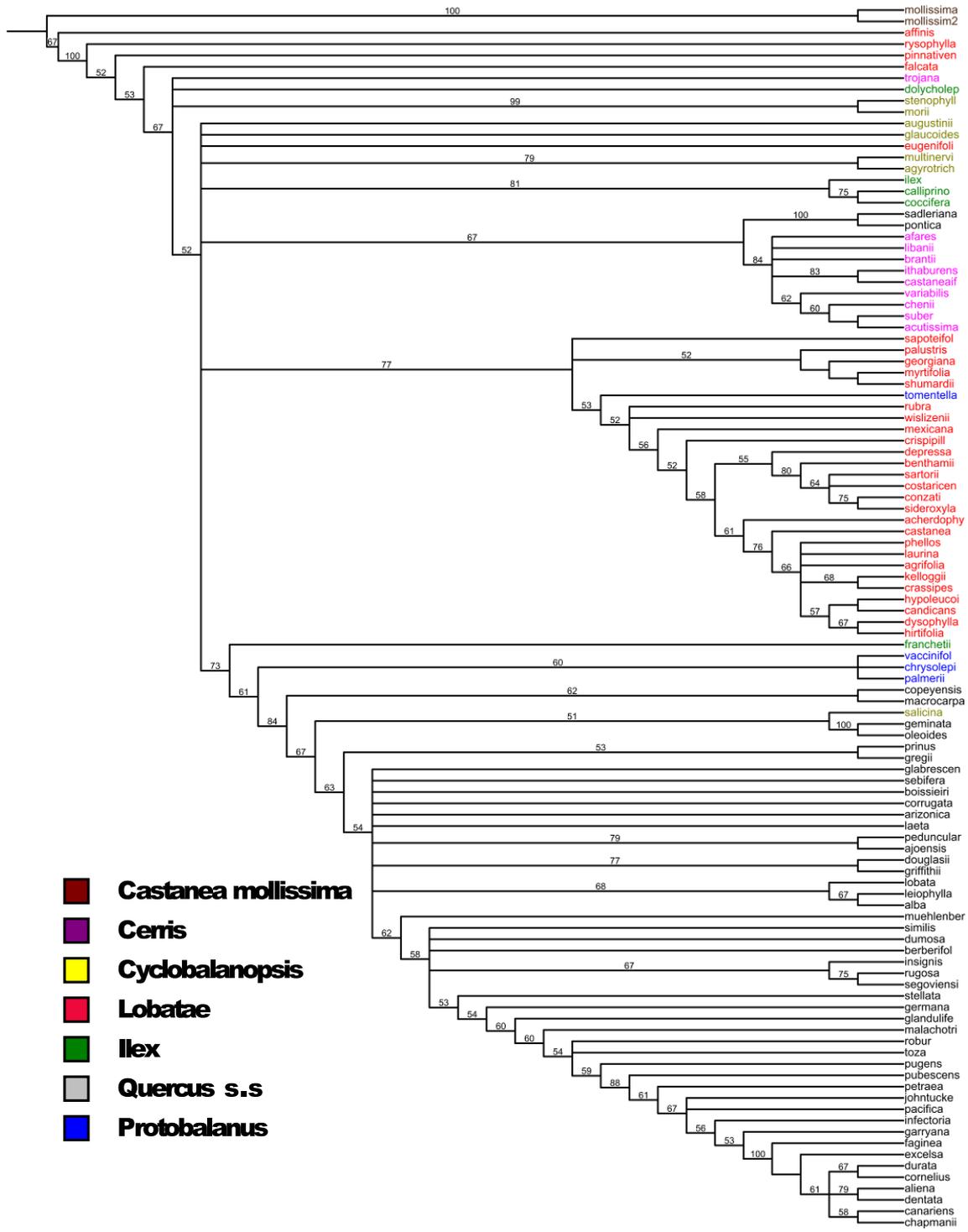


FIGURE B7 : REPRESENTATION PAR SUPERARBRE DES 4 COS SANS LONGUEUR DE BRANCHE, SEULS LES BOOTSTRAPS SUPERIEURS A 50% SONT INDIQUES.

Plusieurs conclusions et questions peuvent être évoquées sur la pertinence de l'utilisation de ces gènes en phylogénie.

Nous pensons que du fait de la persistance de ces gènes au sein d'espèces très différentes, ces différents gènes pourraient avoir gardé un pouvoir de résolution sur les nœuds anciens. Or, nous voyons qu'en regardant de façon différente ces gènes, la concaténation semble être la façon la plus intéressante pour analyser ces données, mais ne permet pas d'obtenir de relations évolutives entre les groupes.

Nous pouvons nous poser la question de la véritable orthologie de ces gènes. En effet, l'étude bioinformatique sur les espèces dont le génome a été entièrement séquencé peut trancher en faveur de l'orthologie. Par contre, l'orthologie de ces gènes chez les trois genres *Quercus*, *Castanea* et *Fagus*, pour lesquels seules des séquences EST étaient disponibles, est beaucoup moins certaine.

Cependant, le fait que l'arbre phylogénétique qui semble le plus pertinent provienne de la concaténation de toutes les séquences des gènes et non pas d'une analyse SuperArbre nous incite très fortement à penser que le nombre de gènes est important dans ce jeu de données et que l'irrésolution que nous voyons à tous les niveaux provienne d'un manque de signal phylogénétique au sein de chaque gène. Ce déficit pourrait être compensé par la prise en compte d'un nombre plus important de gènes, au delà des quatre que nous avons analysés et/ou par la prise en compte d'un plus grand nombre de taxa.

Dans notre cas et vu que les principaux groupes sont représentés avec un nombre de taxa assez élevé (110), nous pensons que le problème viendrait donc principalement du faible nombre de gènes utilisés et/ou de l'information portée par l'histoire de ces gènes.

A ce stade et en utilisant uniquement quatre gènes présents dans des espèces éloignées et constituant des COS, la reconstruction phylogénétique n'a pas permis d'aboutir à un arbre phylogénétique résolu, la raison la plus vraisemblable étant un manque d'information contenue dans ces gènes. L'utilisation d'une autre source d'information s'avère donc nécessaire, c'est dans ce cadre que l'utilisation de gènes issus du projet de reséquençage allélique s'avère particulièrement utile et fait l'objet de la partie suivante.

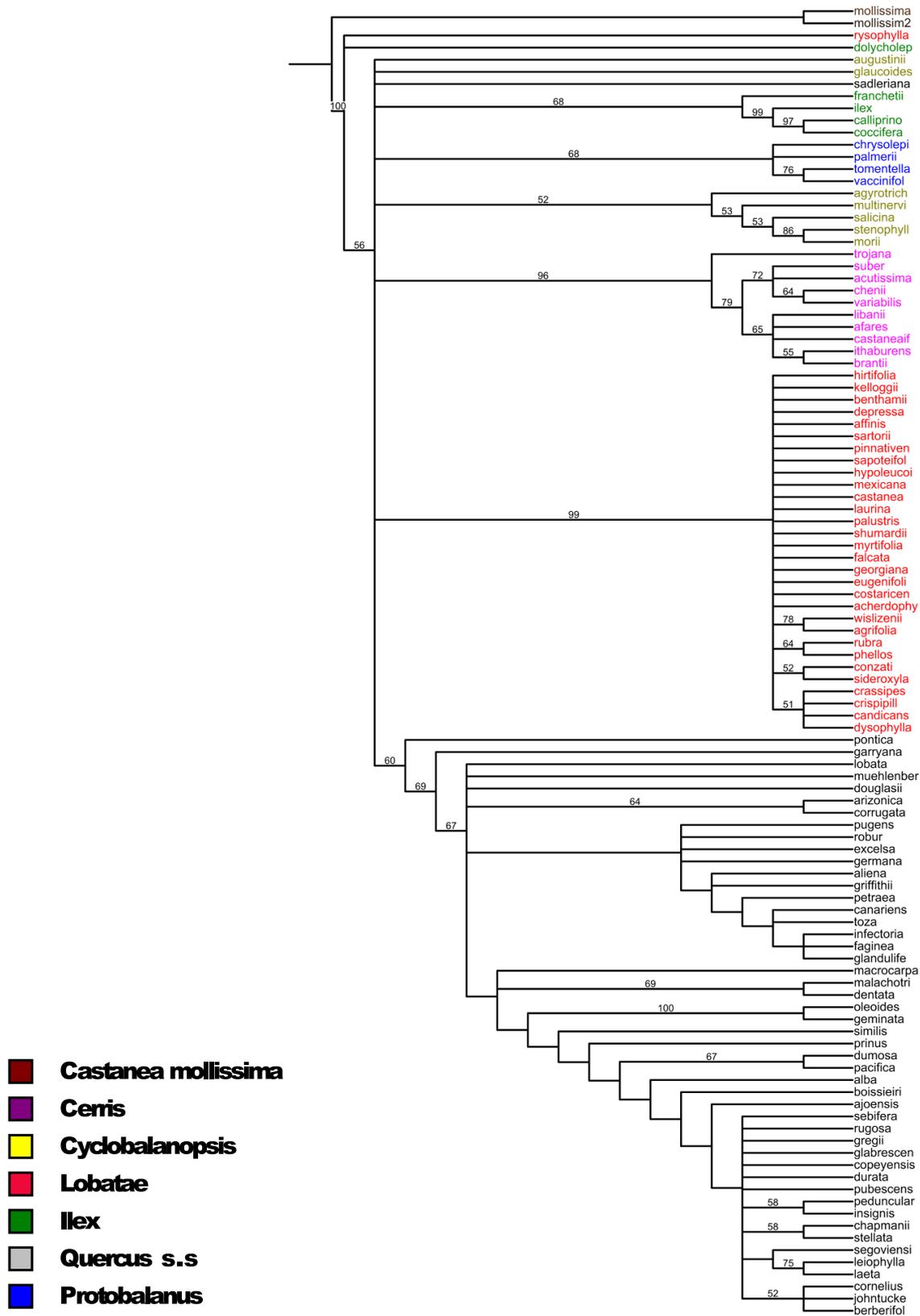


FIGURE B8 : REPRESENTATION PAR SUPERMATRICE DES 4 COS SANS LONGUEUR DE BRANCHE, SEULS LES BOOTSTRAPS SUPERIEURS A 50% SONT INDIQUEES.

La recherche de marqueurs moléculaires de type COS a été le premier travail que j'ai entrepris durant ma thèse. J'ai profité à cette occasion de l'expertise d'un post doctorant (Nicolas Feau) qui avait déjà travaillé en phylogénie et connaissait assez bien ce que rechercher de nouveaux gènes avait comme implications.

Cette recherche à été assez aisée dans le sens où, sur les 27 gènes que j'ai testés en PCR, l'amplification était soit très bonne, soit très mauvaise ; je n'ai pas donc pas eu à adapter ni les protocoles d'extractions ni les protocoles d'amplifications. La partie séquençage s'est également bien déroulée.

Par contre, une fois les quatre gènes entièrement amplifiés sur tous les individus de notre échantillonnage, une part très importante de mon travail a été de comprendre pourquoi on ne trouvait pas de signal.

Il est assez aisé de tester une seule méthode quand les résultats sont satisfaisants, mais quand ils ne le sont pas, il est nécessaire d'emprunter différentes voies afin de ne pas passer à côté d'un paramètre fondamental. C'est dans ce cadre là que je me suis familiarisé avec le logiciel PhyML ainsi qu'avec ses nombreux paramètres. Je me suis également intéressé aux différents logiciels capables d'inférer des SuperArbres, mais à chaque fois le résultat était assez décevant. Le fait de devoir rechercher des solutions techniques à ce genre de problème a été plus que motivant pour moi dans le sens où je pouvais laisser s'exprimer le côté « Ingénieur » de ce genre de travail.

Je n'ai donc pas été déçu par le fait qu'à ce stade, le travail en laboratoire et bioinformatique sur les COS n'ait pas donné de résultats satisfaisants dans le sens ou comme dans les études précédentes, il est possible de retrouver les groupes taxonomiques mais une irrésolution apparaît aux nœuds profonds.

La recherche du second type de marqueur nucléaire s'est faite de façon parallèle à celle que l'on vient de voir, de telle sorte qu'en Octobre 2012, tous les gènes étaient disponibles sur toutes les espèces. A cette date, j'ai participé en tant qu'orateur à un colloque International (IUFRO) organisé par mon laboratoire d'accueil ou j'ai présenté mes travaux. Avec l'aide de mon directeur de thèse (Antoine Kremer) et de Vincent Berry qui est spécialiste des méthodes SuperArbres, nous avons pris la décision d'exclure les COS des analyses suivantes à cause de leur manque de signal mais également à cause du fait qu'ils perturbent le signal des gènes du reséquençage lorsqu'ils sont mis ensemble dans une SuperMatrice (données non montrées).



## B2.8 REFERENCES BIBLIOGRAPHIQUES

---

- Denk, T. and G. W. Grimm (2010). "The oaks of western Eurasia: Traditional classifications and evidence from two nuclear markers." Taxon **59**(2): 351-366.
- Gouy, M., S. Guindon, et al. (2010). "SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building." Molecular Biology and Evolution **27**(2): 221-224.
- Guindon, S. and O. Gascuel (2003). "A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood." Systematic Biology **52**(5): 696-704.
- Hall, T. A. (1999). "BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT." Nucleic Acids Symposium Series **41**: 95-98.
- Mayer, K., C. Schuller, et al. (1999). "Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*." Nature **402**(6763): 769-777.
- Posada, D. and K. A. Crandall (1998). "MODELTEST: testing the model of DNA substitution." Bioinformatics **14**(9): 817-818.
- Ronquist, F. and J. P. Huelsenbeck (2003). "MrBayes 3: Bayesian phylogenetic inference under mixed models." Bioinformatics **19**(12): 1572-1574.
- Samuel, R., A. Bachmair, et al. (1998). "ITS sequences from nuclear rDNA suggest unexpected phylogenetic relationships between Euro-Mediterranean, East Asiatic and North American taxa of *Quercus* (Fagaceae)." Plant Systematics and Evolution **211**(1-2): 129-139.
- Sato, S., Y. Nakamura, et al. (2000). "Structural Analysis of *Arabidopsis thaliana* Chromosome 5. X. Sequence Features of the Regions of 3,076,755 bp Covered by Sixty P1 and TAC Clones." DNA Research **7**(1): 31-63.
- Stamatakis, A. (2006). "RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models." Bioinformatics **22**(21): 2688-2690.
- Ueno, S., G. Le Provost, et al. (2010). "Bioinformatic analysis of ESTs collected by Sanger and pyrosequencing methods for a keystone forest tree species: oak." Bmc Genomics **11**: 650.
- Xu, L. (2004). Diversité de l'ADN chloroplastique et relations phylogénétiques au sein des Fagacées et du genre *Quercus*. Thèse de l'Université Henri Poincaré, Nancy, France.



---

---

B3 PHYLOGENETIC RELATIONSHIPS BETWEEN INFRAGENERIC  
GROUPS IN THE GENUS QUERCUS

---

---

François Hubert, Emmanuelle Jousselin, Vincent Berry, Guido W. Grimm, Alain Franc, Antoine  
Kremer



La présente partie est écrite sous la forme d'un article en anglais qui sera soumis dans les semaines à venir à la revue « International Journal of Plant Biology ».

Cet article est basé sur des données moléculaires à partir de six gènes nucléaires issus d'un projet de reséquençage allélique qui a débuté avant cette thèse. Les espèces utilisées (112 au total) sont les mêmes que dans la partie précédente et je renvoie le lecteur à la partie Matériel et Méthodes (B1) s'il en ressent le besoin.

A la suite de l'article, une partie en français expliquera ce qu'il faut retenir de cette étude de façon très simple, afin que la lecture des trois parties (marqueurs COS, marqueurs nucléaires de la présente partie, et séquençage du génome chloroplastique) puisse se faire aisément et que le message principal passe bien. Nous ajouterons également le résultat de la phylogénie calibrée avec les 5 datations issues de l'analyse de la partie A4. Cette figure sera peu discutée étant donné qu'elle va servir de base pour la poursuite du projet (le projet PHYLOSPACE se terminant fin 2013).

Tableau B5. Number of species collected and analyzed.

Genus	Subgenus	Groups	Geographic origin	No. Species f	Estimated number of species
<i>Castanea</i>			Asia	1	
<i>Notholithocarpus</i>			America	1	
<i>Quercus</i>					
	<i>Cyclobalanopsis</i>		Asia	9	87e, 76c
	<i>Quercus s.l.</i>				
		<i>Quercus s.s.</i>	Asia	6	25e, 15c
			Europe and North Africa	8	16d
			America	35	104a
		<i>Lobatae</i>	America	31	82a
		<i>Protobalanus</i>	America	4	5b
		<i>Cerris</i>	Asia	6	8c
			Europe and North Africa	4	4d
		<i>Ilex</i>	Asia	3	26c
			Europe and North Africa	2	2d
				109	

- a) Nixon K.C., 1993b.  
 b) Floras (Brach and Song, 2006), published on the Internet <http://www.efloras.org> ).  
 c) Menitsky Y.L. 2005.  
 d) Flora europea published on the internet (<http://rbg-web2.rbge.org.uk/FE/fe.html>).  
 e) Camus (1936-1954).  
 f) Assignment of the study species to infrageneric groups *Cyclobalanopsis*, *Quercus s.s.*, *Lobatae* and *Protobalanus* was done according to Camus (1936-1954) and Valencia (Valencia, 2004). Assignments to *Cerris* and *Ilex* was done according to Menitsky Y.L. 2005

### B3.1 INTRODDUCTION

---

Oaks (*Quercus* spp) belong to one of the most widely distributed forest-tree genus. Besides their economic and ecologic importance, oaks are also considered in many countries as cultural and patrimonial resources. Depending on the authors, there are between 300 and 600 reported oak species. Since the beginning of the Linnaean taxonomy, the classification within the genus *Quercus* has been controversial, and more than 20 classifications have been proposed see Denk and Grimm (2010) for a review of some of the classifications. Indeed species delineation in oaks is hampered by the very large intraspecific genetic variation reported in the genus (Kremer and Petit, 1993; Kremer *et al.*, 2012) and frequent interspecific hybridization (Rushton, 1993). During the past decade, phylogenetic investigations based on traditional molecular markers have clarified the infrageneric classification by recognizing six major groups (Manos *et al.*, 2001; Oh and Manos, 2008; Denk and Grimm, 2010) which are also partially supported by classification systems of Camus (1936-1954), Schwarz (1936) and Menitsky (2005). Morphological traits are highly homoplastic, but five of the six groups can be unambiguously identified based on pollen ornamentation and ultrastructure (Denk and Grimm, 2009; Denk *et al.*, 2012). However earlier phylogenetic reconstructions did not solve the relationship among these groups due to either taxon sampling, (these groups were incompletely represented), gene sampling and outgroup selection.

Here we propose to explore the recent developed genomic resources within the genus (Ueno *et al.*, 2010) to find additional nuclear gene sequences for clarifying the phylogenetic relationships among the six groups, by using a comprehensive sampling of taxons in each of the six infrageneric groups and across their biogeographic distribution.

Different infrageneric classifications have been proposed and the taxonomic ranks given to groups of species differed depending on authors. Here, we treat oaks as one genus with two subgenera (*Cyclobalanopsis* and *Quercus*) as earlier done by Ørsted (Ørsted, 1867), Rehder (Rehder, 1940) Camus (Camus, 1936-1954) and adopted the subdivision of subgenus *Quercus* into five major infrasubgeneric groups as suggested by Nixon (Nixon, 1993): the white oaks (*Quercus*), the red and black oaks (*Lobatae*), the intermediate oaks (*Protobalanus*), the *Cerris* group and the *Ilex* group. These groups have received different ranks either as subgenera or sections by different authors. While there is a consensus about species assignments to the three first groups, there are some disagreements concerning the delimitation of the two latter groups. We adopted the classification of Menitsky (Menitsky, 2005) for the *Cerris*, *Ilex* and white oak groups (*Cerris* group = Subgenus *Quercus*, section *Cerris*; *Ilex* Group= Subgenus



*Heterobalanus*; white oaks = Subgenus *Quercus*, section *Quercus*). Menitsky's treatment corresponds to the most recent revision of the Eurasian oaks. The correspondence of Menitsky's classification with others is reviewed in Denk and Grimm (Denk and Grimm, 2010). Altogether we will therefore consider six taxonomic groups by adding subgenus *Cyclobalanopsis* to the five groups in the subgenus *Quercus*. To avoid any further confusion, we will designate by *Quercus sensu lato* (= *Quercus s.l.*) the subgenus *Quercus* and by *Quercus sensu stricto* (= *Quercus s.s.*) the infrageneric group within *Quercus s.l.* In previous phylogenetic analyses these six species groups were not always recovered as clades (Manos *et al.*, 1999; Manos and Stanford, 2001; Manos *et al.*, 2001; Oh and Manos, 2008; Denk and Grimm, 2009, 2010). The phylogenetic relationships between these clades were not clear, and reconstructions obtained so far have given contrasted results. When phylogenetic analyses were based on chloroplast DNA sequences, the six groups formed a polytomy (Manos *et al.*, 1999; Xu, 2004). As more nuclear sequences were added to the analyses (ITS or Crabs claw for example) the six groups were dispatched into two "super clades": one clade gathered the subgenus *Cyclobalanopsis* and both infrageneric groups *Cerris* and *Ilex*, and the second clade gathered the remaining three groups *Protobalanus*, *Lobatae* and *Quercus* (Oh and Manos, 2008; Denk and Grimm, 2010). Interestingly the first clade is only present in Eurasia, while the second extends mainly in America, with group *Quercus s.s.* however being present also in Asia and Europe. This topology suggests that subgenus *Quercus s.l.* is diphyetic as subgenus *Cyclobalanopsis* is nested within the Eurasian superclade of subgenus *Quercus s.l.*. Within each "super clade", the topology varies depending on the studies. For instance, in the New World superclade, the *Protobalanus* group is either found as a sister clade to the *Quercus s.s.* group (Manos *et al.*, 1999), or as a sister clade to the *Lobatae* and *Quercus s.s.* groups. Within the Eurasian superclade, there are also alternative topologies, as subgenus *Cyclobalanopsis* may either be excluded (Manos *et al.*, 1999; Denk and Grimm, 2009) or included (Oh and Manos, 2008; Denk and Grimm, 2010). In this study in order to stabilize phylogenetic relationships among traditionally recognized taxonomic groups in the genus *Quercus*, we explored a wider set of new unpublished nuclear genes taking advantage of recent developments of genomic resources in the Fagaceae. Our objectives were threefold (1) to test the monophyly of all of the infrageneric groups, (2) to investigate the consistency of the relationships within the two superclades that have been unraveled by previous reconstructions, (3) to improve the resolution at the deeper nodes of the phylogeny.



---

## B3.2 MATERIAL AND METHODS

---

---

### B3.2.1 SPECIES SAMPLING

---

Leaf or bud material for 108 species of the genus *Quercus*, and two additional species belonging to related genera *Notholithocarpus* and *Castanea* (Tableau B5) were collected. All samples were collected *in situ* from natural populations or collections (mostly common garden experiments). Within the genus *Quercus*, nine species of subgenus *Cyclobalanopsis* and 99 species of subgenus *Quercus s.l.* were sampled.

Leaf or bud material was collected on one to seven individuals for each species located within a neighbourhood of less than 500 meters. The overall collection of specimens (210 individuals) comprised a comprehensive and representative diversity of species of the major infrageneric groups identified so far (*Cyclobalanopsis*, *Quercus s.s.*, *Lobatae*, *Protobalanus*, *Cerris* and *Ilex*) across their geographic distribution (Tableau B5). A complete list of samples, their taxonomic classification and their geographic origin can be found in Annexe 2.

---

### B3.2.2 DNA EXTRACTION, MARKER SELECTION, AMPLIFICATION AND SEQUENCING

---

Collected tissues were immediately preserved in dried silicagel or were frozen at -80°C. Genomic DNA was extracted using the Invisorb DNA plant HTS 96 kit from Invitex Company using the manufacturer's protocol. DNA quality was visualized on a 1 % (w/v) agarose gel stained with GelRed (Biotium, USA). DNA concentration was quantified on an eight channel Nanodrop spectrophotometer and concentration of each sample was adjusted manually to approximately 10 ng/ µL.

We explored existing genomic resources available in the *Quercus* subgenus in order to identify multiple housekeeping genes that might be useful for phylogenetic reconstructions.

Tableau B6: List of the primers used for PCR amplifications

Sequence name	Forward (F) / Reverse (R)	Sequence
1155	F	GCTTGAGCCTTCCGAAAGTG
	R	CGCATATTGCATCTCTCAACTCATC
8461	F	CGCTGCTTATGGTCATTTCCGG
	R	GAGTCATCATAGACAATGCAGGTGG
8561	F	CACCACAATAAAGGGC
	R	CCAATTCCTCAAAACCTC
5191	F	AAAGGTGGCTGAGATTTTG
	R	AAAACCCGCATTAGTGTTTC
9715	F	TTGCTGTGYGGATGAGAAG
	R	CCAAAACCAACCACAATAAC
8745	F	CAGGATGGCATATTGTTTG
	R	TGATTGTCCGATGGCTC

We used the resequencing data of a set of 800 markers in *Q. petraea*, *Q. robur* and *Q. ilex*. This resequencing project of 800 markers was conducted on 11 *Q. petraea*, 11 *Q. robur* and 1 *Q. ilex*. These 800 markers were made of 612 candidate genes (<http://www.evoltree.eu/index.php/candidate-genes>) potentially involved in adaptive traits (such as growth, bud phenology, biotic interactions...) and a subset of 87 control genes randomly selected in the EST database of *Quercus petraea* and *Q. robur* (Ueno *et al.*, 2010). These 87 genes were analyzed to select a subset of markers exhibiting interspecific variation between *Q. robur*, *Q. petraea* and *Q. ilex* and no intraspecific variation. We set several criteria in order to select genes that might be useful used for phylogenetic investigations. To ensure that they could be sequenced in one run, we set length limits to 500 and 900 bp. We also took into account divergence between *Q. petraea/Q. robur* and *Q. ilex*. Finally amplification success and sequence quality were used as final decisive criteria. Using these criteria, we selected six markers annotated throughout the Blastx searches on the Swissprot database (release May 2012) with an E-value cutoff of 10<sup>-5</sup> (CL1155: No hit, CL8461: No hit, CL8561: NADP-dependent malic enzyme, CL5191: 1-aminocyclopropane-1-carboxylate oxidase homolog 1, CL9715: No hit, CL8745: Calnexin homolog). The primers sequences are listed in Tableau B6. All these preliminary tests aiming at the selection of genes were conducted on a panel of 14 individuals, comprising two individuals from each of the six infrageneric groups of the genus *Quercus* and 2 specimens of *Castanea*.

All amplifications were carried out in a 20 µl volume, containing 2µl of 10x reaction Buffer, 5mM of MgCL<sub>2</sub>, 5 mM each dNTP, 4 µM each primer, 0,5 U of Amplitaq Gold(R) 360 (Applied biosystems) and approximately 20 ng of template DNA. The same Touch Down PCR programs were used for all the PCR amplification as follows :

An initial step at 95°C for 10 min, followed by 2 cycles at 94°C for 60, 59°C for 60 s, 70°C for 120 s, 17 cycles at 93°C for 45 s, 59°C for 30s with a decrease of temperature of 0,5°C per cycle, 70°C at 120 s, 19 cycles at 92°C for 30 s, 50°C at 30 s, 70°C at 120 s and a final incubation at 72°C for 10 min. PCR products were separated on 3 % agarose gel stained with GelRED (Biotium, USA) for verification of amplification. All amplified fragments were sequenced in an ABI3730 automated sequencer. The obtained sequences were imported into Codon Code Aligner version 3.7 (CodonCode Corporation, Dedham, MA, USA). We only used sequences where forward and reverse reads were of good quality. All other sequences that didn't match these criteria were discarded.

Tableau B7. Nuclear gene regions features for the 6 genes studied and CRC and ITS regions.

Gene name	Gene annotation	Length (bp)	Number of species		Number of variable characters / lenght		Number of parsimony informative characters / lenght		Number of distinct patterns		Chosen models
			All samples	Quercus only	All samples	Quercus only	All samples	Quercus only	All samples	Quercus only	
CL1155	No hit	550	83	79	0,2	0,19	0,15	0,14	124	126	HKY
CL8461	No hit	210	65	61	0,15	0,13	0,09	0,086	53	57	HKY+ $\Gamma$
CL8561	NADP-dependent malic enzyme	653	79	76	0,21	0,17	0,16	0,11	155	161	HKY + $\Gamma$
CL5191	1-aminocyclopropane-1-carboxylate oxidase homolog 1	529	48	46	0,17	0,14	0,14	0,12	113	114	HKY+ $\Gamma$
CL9715	No hit	389	93	89	0,19	0,17	0,13	0,11	96	98	HKY+I
CL8745	Calnexin homolog	565	86	86	0,16	0,16	0,12	0,12	95	94	GTR+ $\Gamma$
Total length (bp)		2896	112	110	0,19	0,16	0,14	0,11			
CRC*		2957	52		0,32		0,18				HKY+ $\Gamma$
ITS*		687	41		0,34		0,25				

\* Data based on the entire family (Oh and Manos, 2008).

All sequences were edited using CodonCode Aligner version 3.7 and aligned with the Muscle algorithm of the Seaview software version 4 (Gouy *et al.*, 2010). All coding parts of the markers were checked for protein coding frameshifts, to eliminate pseudogenes (Zhang and Hewitt, 1996) with Mega v4.1 (Tamura *et al.*, 2007). Insertion–deletion events were treated as missing data.

When sequences from several individuals were available for a given species, we compiled a unique consensus sequence by using Bioedit v7.1.3 (Hall, 1999). This strategy allowed to minimize missing data (i.e. number of taxa for which gene sequences were missing). When for a site a nucleotidic base was less frequent than 95%, the IUPAC (International Union of Pure and Applied Chemistry) ambiguity code was used. Accession numbers for each species are provided in Annexe 3.

For the combined analysis we concatenated all sequences into a single alignment using Seaview software version 4. In total we analyzed 112 taxons and six genes representing in total 2896 bp.

---

### B3.2.3 PHYLOGENETIC RECONSTRUCTIONS

---

Phylogenetic trees were rooted using *Notholithocarpus* or *Castanea* samples as outgroups. Recent phylogenetic analysis of Fagaceae showed that these two genera were the closest related relatives of the genus *Quercus* (Manos *et al.*, 2008; Oh and Manos, 2008).

We first performed phylogenetic analyses on each gene using Maximum Likelihood (ML) optimization. ML optimization and Bayesian inferences (BI) were then performed on the matrix of the concatenated sequences.

ML phylogenetic analyses were conducted on each individual gene with Phym1 3.0 (Guindon and Gascuel, 2003). First, the computer program Modeltest version 3.7 (Posada and Crandall, 1998) was used to select an appropriate substitution model for each DNA fragment (Tableau B7). ML trees were inferred under the AIC-selected (Akaike Information Criteria) substitution model. Branch support was established using non parametric bootstrapping and 100 bootstrap replicates. No strong phylogenetic signal was found in any of each marker, but the lack of conflictual topologies allowed us to concatenate these six sequences to conduct analyses on the combined matrix. This matrix was subdivided into six DNA partitions corresponding to each of the six genes.



We performed ML analyses on the combined analyses and associated bootstrapping using the MPI-parallelized RAxML 7.2.8-ALPHA (Stamatakis *et al.*, 2008). Given that  $\alpha$ , which scales the shape of the gamma distribution, and the proportion of invariable sites cannot be optimized independently from each other and following Stamatakis' personal recommendations (RAxML manual), we used GTR +  $\Gamma$  with 4 discrete rate categories for all partitions (-m GTRGAMMA). Support was established based on 1000 bootstrap replicates.

For BI analysis (using MrBayes version 3.1.2), the best fitting model was set for each partition. Parameters of the model were treated as unknown variables with uniform prior probabilities and were estimated during the analysis; they were allowed to vary across partitions. Each Bayesian analysis included two MCMC runs, each composed of four chains, three heated and one cold. Each Markov chain started from a random tree and run for up to  $20 \times 10^6$  generations, sampling the chains every 1000th cycle. At the end of each run we considered the sampling of the posterior distribution to be adequate if the average standard deviation of split frequencies was  $< 0.05$ . The log-likelihood scores of sample points were plotted against generation time to determine when the chain became stationary. After discarding the "burn-in" samples (25% of the trees) MCMC runs were summarized and further investigated for convergence of all parameters, using *sump* and *sumt* commands in MrBayes version 3.1.2 and the computer program Tracer version 1.4 (Rambaut and Drummond, 2007). Data remaining after discarding burn-in samples were used to generate a majority-rule consensus tree where the percentage of samples recovering any particular branch of the consensus tree represented the clade's posterior probability (Huelsenbeck and Ronquist, 2001). Posterior probabilities (PP) of 95% or higher were considered significant support. The mean, variance, and 95% credibility intervals were calculated from the set of substitution parameters.

RAxML and MrBayes analyses were conducted on a 150 cores Linux Cluster at CBGP (Center for Biology and Management of Populations, Montferrier-sur-Lez, France) and a 16 cores Linux Cluster at INRA Bordeaux, France.

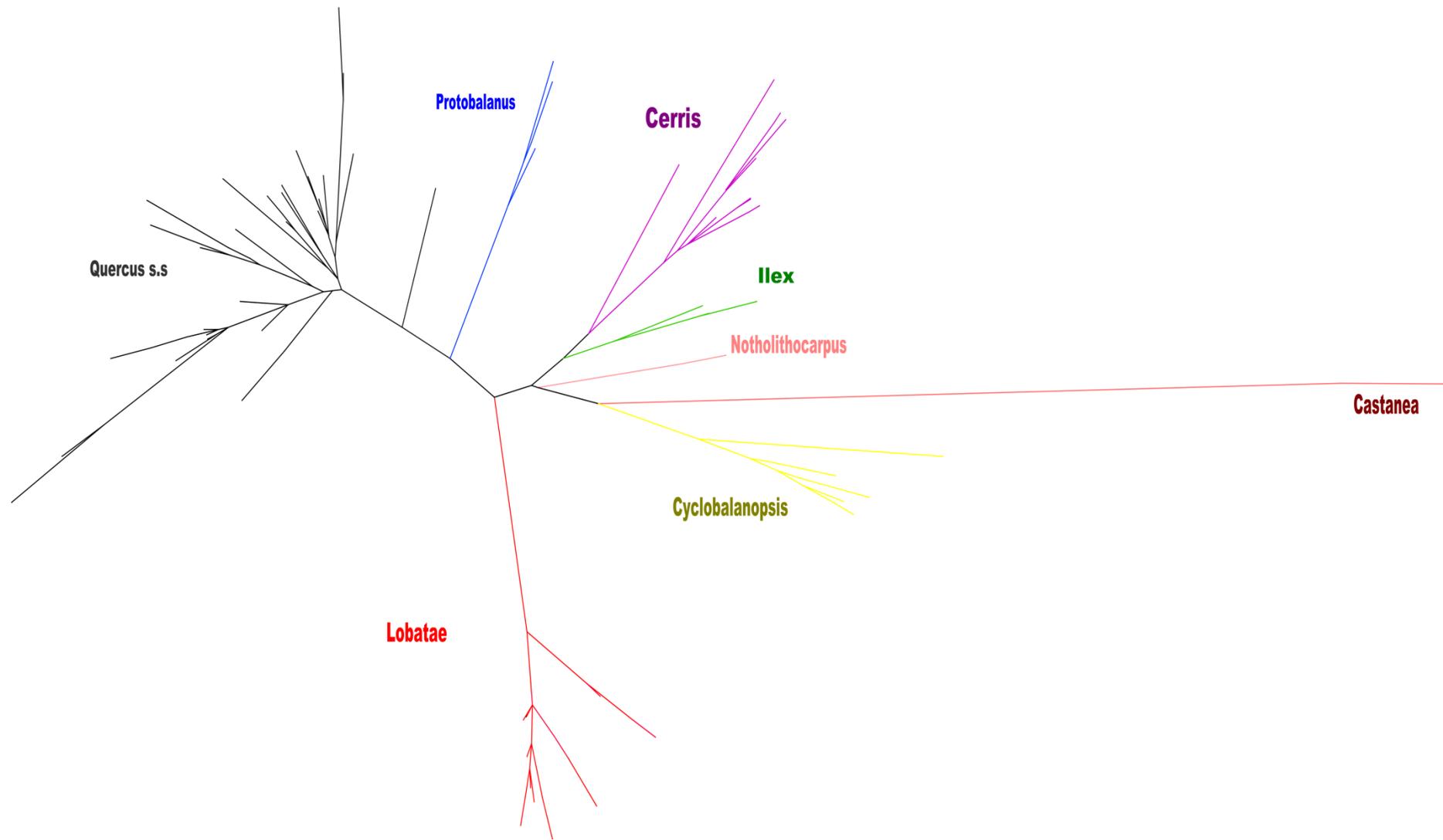


FIGURE B9: UNROOTED TREE OF THE GENUSES QUERCUS, CASTANEA AND NOTHOLITHOCARPUS BASED ON THE 6 CONCATENATED REGIONS (2896 PB).

---

## B3.3 RESULTS

---

---

### B3.3.1 PCR AND SEQUENCING RESULTS

---

Annexe 2 provides a summary matrix of the gene sequences sequences/species that were finally used for the phylogenetic reconstructions. Amplifications were successful for most but not all six nuclear genes selected (Tableau B7). Twelve species were represented by only two gene sequences, 15 by all six gene sequences. The whole dataset consisted in a supermatrix with 446 sequences out of the 672 possible sequences (=112 species \* 6 genes), thus representing 66 % of the theoretical possible data. Overall the six selected gene sequences exhibited less nucleotide diversity and parsimonious informative characters than previous markers used for phylogenetic reconstruction (Tableau B7). The number of distinct alignment patterns varied from 57 to 161 according to the selected genes and including the in and outgroup (Tableau B7). Finally the proportion of ambiguous nucleotidic sites, which comprise the intra species polymorphic sites and the possible sequencing errors, amount to 0.0024, which is rather low in comparison to the between species variation that is shown on Tableau B7.

---

### B3.3.2 SINGLE GENE ANALYSIS

---

In the ML analysis conducted independently on each gene, only the nodes supported by bootstrap proportion values (BP) superior to 50% are represented (Annexe 4 and 5). The phylogenies were rooted on *Notholithocarpus*, and *Castanea* except for CL8745 which was unrooted, because this gene could not be amplified in both genera. The phylogenetic analysis of CL1155 alone retrieved five clades with quite high BP supports (BP generally > 90), while *Protobalanus* and *Quercus s.s.* specimens clustered together in another strongly supported clade (Annexe 4A and 5A). Phylogenetic analyses obtained with CL8561 recovered three clades: *Lobatae*, *Quercus s.s.* and *Cyclobalanopsis* group with BP of 90%, 67% and 80% respectively (Annexe 5B), but only two (*Lobatae* + *Quercus s.s.*) when rooted on *Castanea* (Annexe 4B). The third marker, CL5191, with which we had less amplification success, was less informative (Annexe 4C and 5C). It divided the species into two groups, one clade



with a mix of species belonging to the *Cerris* and the *Cyclobalanopsis* and a polytomous grade with the remaining infrageneric groups. In addition there are some odd-placed taxons outside (one red oak in the *Cerris-Cyclobalanopsis* clade, one *Cyclobalanopsis* at the basis of the *Quercus s.s.-Lobatae* clade) and *Ilex* is unresolved in a basal polytomy. Analyses of CL8461 supported the monophyly of the *Protobalanus* group, the *Lobatae* group and the *Cerris* group with one *Ilex* species (Annexe 4D and 5D). Phylogenetic analyses based on CL8745 only retrieved the monophyly of the *Lobatae* group with a BP of 83 % (Annexe 4E and 5E). Phylogenetic analyses obtained with CL9715 retrieved the monophyly of the *Lobatae* group and the *Cerris* group (Annexe 4F and 5F). Overall single gene analysis retrieved one or several terminal infrageneric groups depending on the gene, but did not allow to solve deeper nodes. These results prevented us from using any supertree approach (Scornavacca *et al.*, 2008), which usually draw a super tree combining the analyses of all single genes.

---

### B3.3.3 COMBINED ANALYSES

---

We first constructed an unrooted tree comprising all species and using the concatenated sequences of all six genes (Figure B9). This tree separates clearly the six major infrageneric groups that we had identified at the beginning. Topologies of rooted trees obtained with the ML approach using the supermatrix of the concatenated sequences were highly concordant with those obtained by BI as shown by the congruence between bootstrap values and posterior probabilities (Figure B10 and B11) and were much better resolved than the trees obtained with each individual gene. However there was a slight topological shift depending on the outgroup species used.

When *Castanea* was used as an outgroup (Figure B10), ML analyses and BI inferences suggested that the genus *Quercus* can be divide into two strongly supported clades corresponding to the two subgenera according to Camus (1936-1954). The nine species of subgenus *Cyclobalanopsis* clustered as a sister clade to the remaining taxa that all belong to the subgenus *Quercus s.l.*

Subgenus *Quercus s.l.* was further subdivided into two moderately supported clades: one clade corresponding to the “New World Oaks” (Manos *et al.*, 1999) comprising *Quercus s.s.*, *Lobatae*, *Protobalanus*, and the second clade was made of species belonging to *Cerris* and *Ilex* groups. However this clade is not comparable to the so called “Old World Oaks” which also includes the *Cyclobalanopsis* group. Three infrageneric groups within the “New World Oaks” were



recovered by ML and BI analyses with strong support values, The *Cerris* and *Ilex* clade comprised one subclade composed of European species of *Ilex* (*Q.Ilex*, *Q.coccifera* and *Q.calliprinos*) and a second subclade gathering the remaining species *Cerris* and Asian *Ilex* (*Q.dolicholepis* and *Q. franchetti*) with a poor support value.

When *Notholithocarpus* was used as an outgroup (Figure B11), the positions of the ingroup root and of the *Cyclobalanopsis* clade change resulting in two poorly supported clades corresponding to the “New World Oaks” as above and the “Old World Oaks” (Manos *et al.*, 1999) that include *Cyclobalanopsis* in addition to *Cerris* and *Ilex*. The main change between both analyses is the disappearance of the deeper split between the two subgenera, and the shift of *Cyclobalanopsis* from a sister taxa position to the “Old World Oaks”.

We positioned different placements of the outgroup for testing alternative ingroup roots, as shown on Figure B10 (when *Castanea* is the outgroup) and on Figure B11 (when *Notholithocarpus* is the outgroup) using the non parametric SH tests implemented in Raxml 7.2.8 ALPHA (Stamatakis *et al.*, 2008). When *Castanea* was chosen as the outgroup, the likelihood at all other placements was significantly lower than the initial position, except when the position was on the branch leading to *Cerris+Ilex* and to *Cyclobalanopsis+Cerris+Ilex*. When *Notholithocarpus* was chosen as the outgroup, the likelihood at all other placements were not significantly different than at the initial position, but was lower when the position was on the branch leading to *Cerris* and to *Cyclobalanopsis*. These comparisons suggest that the only compatible placement of the outgroup at which the likelihood is the highest in both scenario is at the *Cyclobalanopsis+Cerris+Ilex* branch.

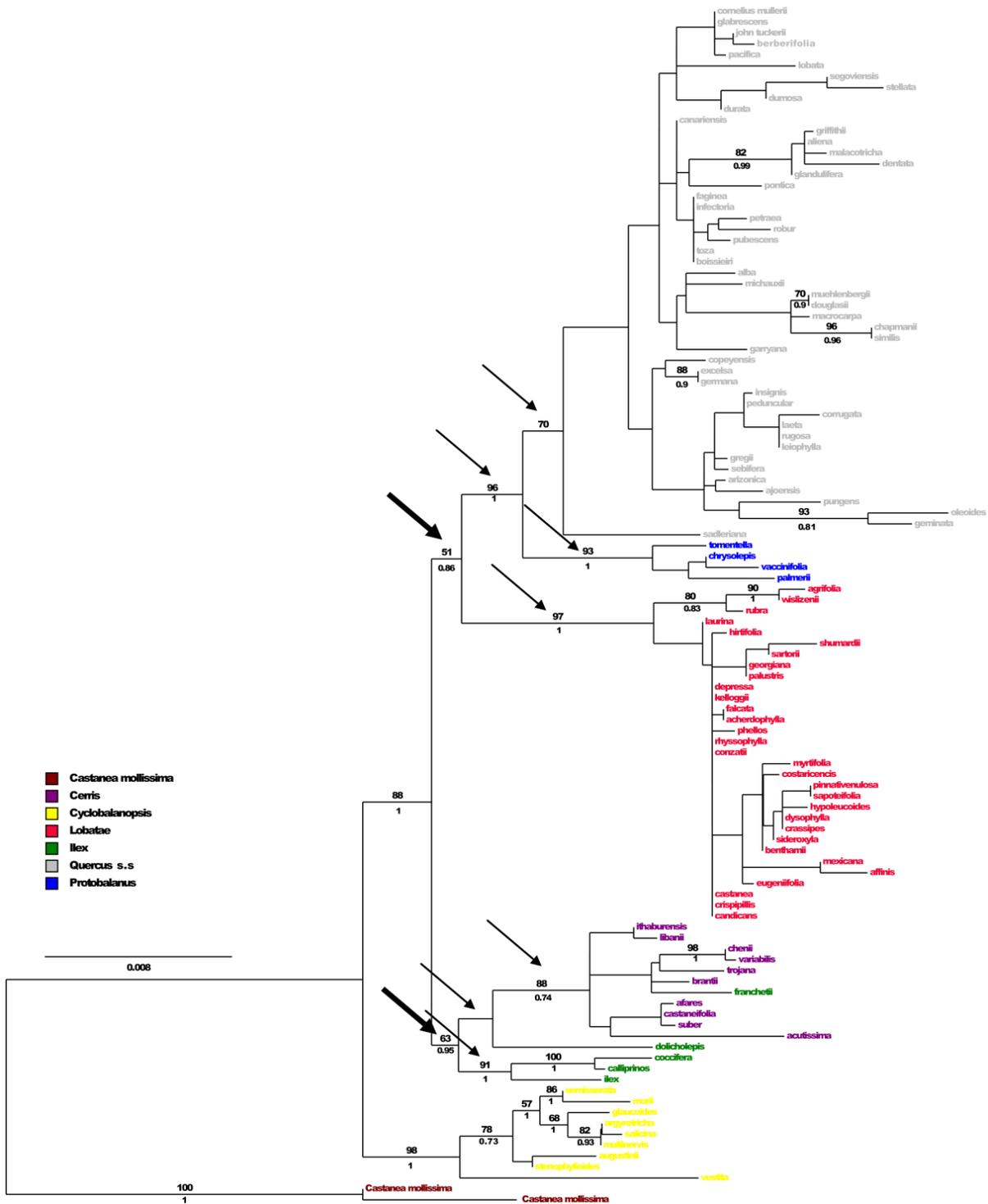


FIGURE B10: MAXIMUM LIKELIHOOD PHYLOGRAM OF THE GENUS QUERCUS BASED ON CONCATENATED SEQUENCES OF 6 GENE REGIONS AND ROOTED ON THE GENUS CASTANEA (2896 BP). NUMBERS CLOSE TO THE NODES REFER TO MAXIMUM LIKELIHOOD BOOTSTRAP SUPPORT VALUES AND POSTERIOR PROBABILITIES. ARROWS INDICATE THE TESTED ALTERNATIVE INGROUP ROOTS.

In this study we explored phylogenetic reconstruction using a larger sample of nuclear genes than in previous investigations and an extensive sample of taxa of the genus *Quercus*. Up to now phylogenetic analyses were based on ITS DNA sequences (Manos *et al.*, 1999; Manos and Stanford, 2001; Manos *et al.*, 2001; Oh and Manos, 2008; Denk and Grimm, 2009, 2010) and of CRABS CLAW (CRC) sequence, a single copy nuclear gene that regulates carpel development in angiosperms (Oh and Manos, 2008). Despite a lower nucleotidic diversity and less parsimony informative characters (Tableau B7) than ITS, our data confirm results from previous studies (Manos *et al.*, 1999; Manos and Stanford, 2001; Manos *et al.*, 2001; Denk and Grimm, 2009, 2010) but also provide new insights into the phylogenetic relationship among the *Quercus* species groups. They provide resolution at deeper nodes but with moderate support, that were continuously observed in earlier investigations conducted with other nuclear markers. They stabilize infrageneric classification within the two major clades, the New and Old World Oaks. However they offer alternative topologies that are dependent on the choice of the outgroup.

---

### B3.4.1 TREE TOPOLOGIES AND OUTGROUP CHOICE

---

Our results lead to two different ingroup roots and, subsequently, two alternative placements of the *Cyclobalanopsis* clade within genus *Quercus*. On the one hand, when *Castanea* was taken as outgroup, our data suggest that the genus *Quercus* comprises two clades corresponding to the subgeneric classification by Camus (Camus, 1936-1954), also accepted by Nixon (Nixon, 1993) and Menitsky (Menitsky, 1984; Menitsky, 2005) : subgenus *Quercus s.l.*, and subgenus *Cyclobalanopsis* (called *Cyclobalanoides* by Menitsky (2005)). This rooted topology differed markedly from the one observed when *Notholithocarpus* was chosen as the outgroup to the tree showing that subgenus *Cyclobalanopsis* joins with *Cerris* and *Ilex* groups to form an Old World clade.

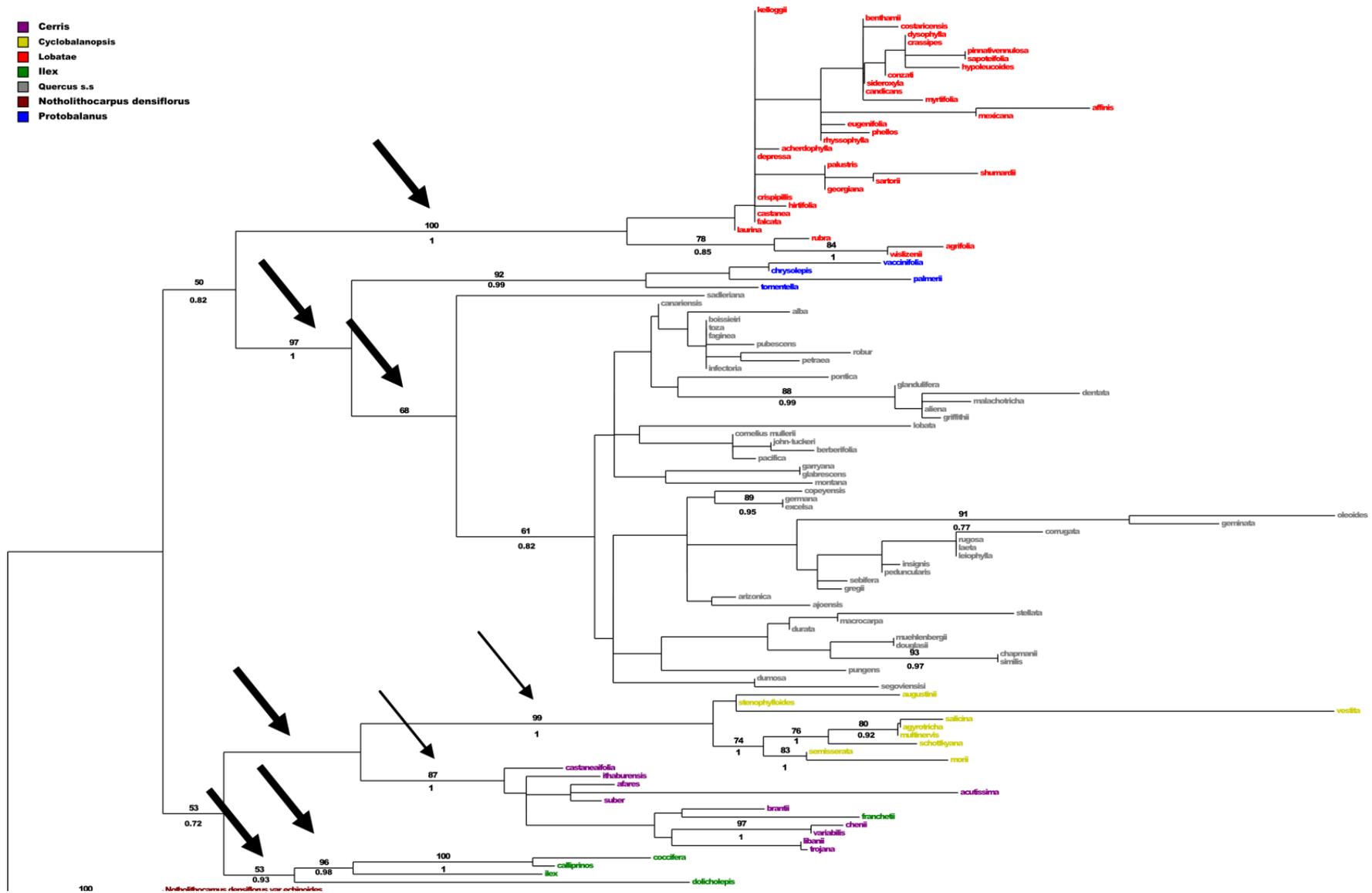


FIGURE B11: MAXIMUM LIKELIHOOD PHYLOGRAM OF THE GENUS QUERCUS BASED ON CONCATENATED SEQUENCES OF SIX GENE REGIONS AND ROOTED ON THE GENUS NOTHOLITHOCARPUS (2896 BP). NUMBERS CLOSE TO THE NODES REFER TO MAXIMUM LIKELIHOOD BOOTSTRAP SUPPORT VALUES AND POSTERIOR PROBABILITIES. ARROWS INDICATE THE TESTED ALTERNATIVE INGROUP ROOTS.

Instability of the genus root was also observed in the past, when different outgroups were used. When distant outgroups were chosen as members of the *Trigonobalanus* genus (Manos *et al.*, 1999), or the *Fagus* genus (Manos *et al.*, 2001), *Cyclobalanopsis* was usually positioned as a sister clade to the other oaks (*Quercus* s.l.). This is a similar topology to the one we observed using *Castanea* as outgroup. However when closely related genera as *Lithocarpus* (Manos and Stanford, 2001) or *Notholithocarpus* (Oh and Manos, 2008) were used as outgroups then *Cyclobalanopsis* was found as member of the Old World Oaks. Phylogenetic relationships between subgenus *Quercus* s.l. and subgenus *Cyclobalanopsis* were therefore ambiguously resolved and are very sensitive to the position and distance of the outgroup. It is very likely that *Castanea* in our data set, being more distant than *Notholithocarpus* from the unrooted network (Figure B9) may attract ingroup taxa that are well differentiated from the rest, such as species belonging to *Cyclobalanopsis*. Thus the split into two subgeneric clades (Figure B10) may actually be an artifact due to long branch attraction (Bergsten, 2005). However one cannot fully exclude that the true phylogeny has actually its root at the longest branch. Solving this issue will require further sampling of taxa within the outgroup to break up long branches. This sensitivity to outgroup choice is a recurrent problem when taxon sampling is incomplete and when long branches are found (De la Torre Barcena, 2009).

---

#### B3.4.2 SUBGENERIC AND INFRAGENERIC SUBDIVISION OF GENUS QUERCUS

---

Within the subgenus *Quercus* s.l., our results confirm infrageneric subdivisions, regardless of the outgroup chosen: the *Cerris-Ilex* clade and the New World clade (*Protobalanus-Quercus* and *Lobatae*) were always retrieved. Within the *Cerris-Ilex* group, our analyses yielded two clades which roughly corresponded to the two botanical groups *Ilex* and *Cerris* (Figure B10 and B11). Resolution is slightly better when *Notholithocarpus* is the outgroup. While one of the Asian *Ilex* species (*Q. franchetii*) is misplaced in both analysis, *Q. dolicholepis* (the second Asian *Ilex* species) is also misplaced when *Castanea* is the outgroup albeit will low support. Reconstructions based on CpDNA sequences (Manos *et al.*, 1999) also separated *Cerris* from *Ilex*, with one Asian *Ilex* species (*Q. phillyraeoides*) misplaced as well within the *Cerris* clade. Evergreen and sclerophyllous European oaks (*Q. ilex*, *Q. coccifera* and *Q. calliprinos*) assigned to the subgenus *Sclerophyllodrys* by Schwarz (Schwarz, 1936-39; Schwarz, 1993) clustered within the *Ilex* clade in our analyses (Figure B10) similarly to previous work based on ITS sequence restricted to Eurasian oak species (Samuel *et al.*, 1998; Bellarosa *et al.*, 2005), and with ITS+5S-IGS data (Denk and Grimm, 2010). However previous studies,



based on CRC sequences and on ITS and CRC sequences suggested that *Cerris* and *Ilex* were mutually monophyletic (Manos and Stanford, 2001; Manos *et al.*, 2001; Oh and Manos, 2008; Denk and Grimm, 2009, 2010), while we found them intermingled in the two groups of the Old World Clade. Because Asian species assigned by Menitsky to the *Ilex* section (*Q. franchetii* and *Q. dolicholepis*) occupied different positions within the Old World clade our data do not support recognition of section *Ilex* (sensu Menitsky) or section *Cerris* (sensu Camus or sensu Menitsky).

Like earlier investigations our results resolve a moderately supported New World clade. However our data are the first to show the subdivision of the New World clade into three subclades (*Protobalanus*, *Lobatae* and *Quercus s.s.*) with strong support. Indeed relative positions of these three groups varied depending on the gene sequences used. With ITS sequences relationships in the New World clade resulted in *Protobalanus* being a sister clade to (*Lobatae* + *Quercus s.s.*), but with poor support (Manos and Stanford, 2001; Manos *et al.*, 2001; Oh and Manos, 2008; Denk and Grimm, 2009). Similar results but with stronger support were obtained with CRC sequences and concatenated CRC+ITS sequences (Oh and Manos, 2008), whereas earlier reports based on ITS sequences and on ITS+CpDNA suggested that *Lobatae* was a sister clade to (*Protobalanus*+ *Quercus s.s.*) (Manos *et al.*, 1999), as we actually found in our results (Figure B10 and B11). In terms of infrageneric classifications, our analysis confirmed the six earlier recognized infrageneric groups *Protobalanus*, *Lobatae* and *Quercus s.s.* groups of the genus *Quercus*: *Cyclobalanopsis*, *Cerris*, and *Ilex*. Interestingly these groups are generally reproductively isolated. Successful natural or control hybridization between species among these groups have only been shown between *Ilex* and *Cerris* (Boavida *et al.*, 2001; Burgarella *et al.*, 2009; Lumaret and Jabbour-Zahab, 2009). Very anecdotal hybridization between *Quercus s.s.* and *Cerris* (Mir *et al.*, 2006) and between *Quercus s.s.* and *Ilex* (Schnitzler *et al.*, 2004) have also been reported. Even if limited, possible hybridization between these groups may explain the taxonomical ubiquitous placement of *Ilex* species either in *Quercus s.s.* by Camus (Camus, 1936-1954), or as a separate group closer to *Cerris* by Menitsky (1984) or Schwarz (Schwarz, 1993). However interspecific hybridization is occurring very frequently within the *Quercus s.s.* group (Rushton, 1993) or the *Lobatae* group (Moran *et al.*, 2012). The possibilities of natural hybridization has been less investigated in the other groups but intra and inter group artificial hybridization experiments support the general conclusion of frequent crossability within groups and very rare hybridization success between groups (Cottam *et al.*, 1982).



---

### B3.4.3 POORLY SUPPORTED PHYLOGENETIC RESOLUTION WITHIN THE GENUS QUERCUS IN DEEPER NODES

---

As earlier phylogenetic reconstructions (Manos and Stanford, 2001; Manos *et al.*, 2001; Oh and Manos, 2008; Denk and Grimm, 2009) our reconstructions are characterized by poorly supported resolution at deeper nodes, a compression of internodes between deeper nodes eg. corresponding to the separation into the New and Old World Oaks, and differentiation into the 6 infrageneric groups. In addition these poorly supported deeper split events produce therefore alternative relationships among the different taxonomic groups. Such patterns are generally considered as signatures of ancient rapid radiation followed by long time spans where later evolution blurred the earlier splits (Rokas *et al.*, 2005). Alternatively similar patterns can also be obtained when the genome has been poorly explored for phylogenetic signals, or when target genes exhibit inappropriate substitution rates to detect ancient events (Whitfield and Kjer, 2008). Before addressing the rapid radiation hypothesis, we may therefore question whether our selected genes were appropriate to solve old speciation events. Selecting genes for detecting targeted phylogenetic signals is a challenging task. On one hand it is difficult to explore the phylogenetic signal before conducting the analysis on the whole data set, on the other hand technical criteria such as PCR amplification success, sequence quality may prevail phylogenetic criteria. Our selection of six protein coding genes resulted from a compromise between technical criteria and optimizing phylogenetic information. But they all yielded congruent footprints of short branches (Annexe 4 and 5). Short internodes were also found in previous reconstructions based on ITS sequences. Congruency of patterns across genes is likely to result from evolutionary processes that impact the whole genome, and therefore would support the rapid radiation hypothesis. Dated splitting of the Old World and New World Clades suggest that it occurred around 17 million years B.P. at the Oligocene–Miocene transition (Manos and Stanford, 2001), although recent fossil data suggest earlier splitting (Denk *et al.*, 2012). This is the time where proliferation of oaks and a burst of diversification in major sections of the genus *Quercus* have been reported throughout the northern hemisphere (Axelrod, 1983). Poor resolution at deeper nodes is no exception to the genus *Quercus*, as it has been repeatedly been found in many plants species (Hughes *et al.*, 2006). The overall conclusion of irresolution at deeper nodes for all genes considered so far further suggests that adding more genes would not necessarily improve the resolution, if no phylogenetic signal is preserved in the genes (Rokas and Carroll, 2005). This hypothesis deserves however to be tested when more markers become available.



### B3.5 BILAN DES RECONSTRUCTIONS PHYLOGENETIQUES BASES SUR DES GENES NUCLEAIRES

---

L'objectif principal qui avait présidé à cette thèse était d'explorer les ressources génomiques disponibles dans le genre *Quercus* pour affiner la phylogénie qui était faite traditionnellement sur la base de l'ITS ou de quelques gènes complémentaires comme le CRC.

Pour ce faire, six nouveaux gènes nucléaires ont été utilisés pour former un arbre phylogénétique. Cet arbre a été réalisé sur 110 individus (espèces), en concaténant les séquences des six gènes (chaque gène séparé donnant peu d'information, la méthode de SuperArbre ne pouvant donc pas être intéressante à ce stade). Le résultat principal obtenu se trouve sur la Figure B11, qui représente l'arbre phylogénétique des 108 espèces de *Quercus*, raciné sur le genre *Notholithocarpus*.

Cet arbre est capable de :

- 1) Retrouver les six groupes taxonomiques définis au début de ce manuscrit (exception faite d'un individu du groupe *Ilex* qui se retrouve dans le groupe des *Cerris*),
- 2) Retrouver les deux clades « New World Oaks » et « Old World Oaks », composé des mêmes groupes que dans les études précédentes,
- 3) Stabiliser la composition du clade des « New World Oaks » : les trois groupes (*Lobatae*, *Quercus s.s* et *Protobalanus*) sont bien soutenus, et il y a association des *Protobalanus* avec les *Quercus s.s* de façon très soutenue, ceci ayant été rarement rapporté dans la littérature,
- 4) Au niveau du clade des « Old World Oaks » : le groupe des *Cyclobalanopsis* est très bien soutenu, la définition des *Cerris* et des *Ilex* semblant être mise à mal.

Mais le point le plus important de cette analyse est que les branches profondes sont également faiblement soutenues, en dépit de l'utilisation de nouveaux gènes (Figure B11). L'hypothèse selon laquelle une radiation rapide aurait touché le compartiment chloroplastique et nucléaire est toujours d'actualité, étant donné que de nouveaux gènes donnent le même pattern.



A ce stade, nous pensons toujours que le manque de soutien aux branches profondes indique un manque de résolution. La définition du groupe des *Cyclobalanopsis* comme sous-genre ne peut être confirmée d'après toutes ces données. La seule certitude provenant de nos données est la définition d'un clade incluant le groupe des *Protobalanus* avec les *Quercus s.s* (Figure B11).

L'ensemble des données nucléaires nous conduit à construire un arbre phylogénétique fictif représentant les hypothèses que l'on a à présent sur la phylogénie du genre *Quercus* (Figure B12). Il s'agit d'un arbre hypothétique qui sera comparé à l'arbre phylogénétique issu du séquençage du génome chloroplastique total :

- Ce genre est séparé en deux sous-genres : le sous-genre *Quercus s.l* et le sous-genre *Cyclobalanopsis* (cette hypothèse provenant des travaux issus de la botanique mais également de notre propres résultats quand *Castanea* est utilisé comme outgroup),
- Les clades « New World Oaks » et « Old World Oaks », mis en évidence par plusieurs études antérieures et nos propres résultats, sont toujours faiblement soutenus,
- Les *Cerris* se retrouvent en groupe frère des *Ilex* (d'après nos données et la majorité des études précédentes),
- Les *Protobalanus* se retrouvent en groupe frère des *Quercus s.s* (d'après nos données uniquement et une seule étude préalable).

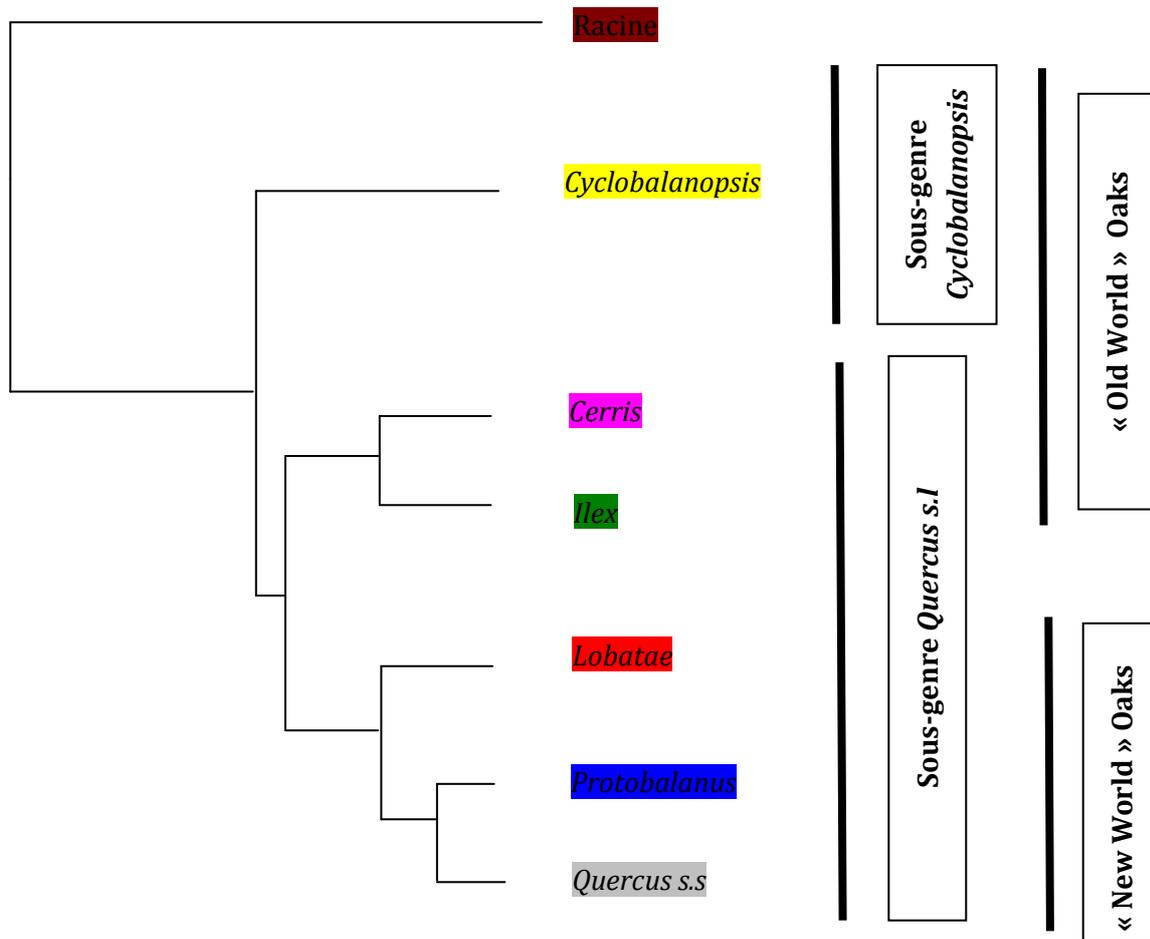


FIGURE B12 : PHYLOGENIE HYPOTHETIQUE BASEE SUR LES DONNEES ISSUES DE LA LITTERATURE (MANOS ET AL., 1999; MANOS AND STANFORD, 2001; MANOS ET AL., 2001; OH AND MANOS, 2008; DENK AND GRIMM, 2010) ET DE NOS PROPRES DONNEES.

### B3.6 PHYLOGENIE CALIBREE BASEE SUR SIX MARQUEURS NUCLEAIRES

---

En prolongement des reconstructions phylogénétiques basées sur les 6 gènes nucléaires et qui font l'objet de l'article en anglais, nous avons tout naturellement essayé de calibrer ces reconstructions dans le temps, en faisant notamment appel au logiciel BEAST (Bayesian Evolutionary Analysis by Sampling Tree) dans sa version 1.7.4 (Drummond *et al.*, 2012), et en valorisant les données fossiles qui ont été accumulées dans la base Fossilmap (chapitre A4.2). Cette échelle est obtenue en modélisant un taux d'évolution moléculaire sur chaque branche de l'arbre. Pour ce faire, les datations fossiles des six groupes (Tableau A13) ont été rentrées dans le logiciel via le logiciel BEAUTI v1.7.4 (Drummond *et al.*, 2012) qui permet de formater le fichier d'entrée pour le logiciel BEAST. Chacun des six gènes est optimisé indépendamment suivant le modèle d'évolution le plus adéquat (Tableau B7), puis, un seul arbre résumant toutes les informations est demandé. La racine utilisée est constituée par le genre *Notholithocarpus*. L'horloge moléculaire recherchée permettant de dater les nœuds est de type relâchée (Drummond *et al.*, 2006), chaque date est modélisée par une loi normale définie par la moyenne des âges extrêmes (Tableau A13) et un écart type englobant celles-ci. Trois cent millions de générations ont été demandées au logiciel, avec un échantillonnage des arbres toutes les 30 000 générations. Le logiciel Tracer v1.4 (Rambaut and Drummond, 2007) a permis de voir si la convergence était atteinte (Tableau B8). En effet, celle-ci l'est si les valeurs des ESS (Effective Sample Size) sont supérieures à 100. Dans ce cas, le nombre de génération a été suffisant pour modéliser convenablement les distributions postérieures des différents paramètres. Afin d'atteindre une valeur d'ESS suffisante (supérieure à 100), il faut multiplier le nombre de générations par la valeur d'ESS la plus faible (9.683 pour le paramètre : gene3.treeLikelihood, Tableau 4) divisée par 100, soit une longueur de chaîne passant de 300 millions à 3 000 millions de générations. Ceci revient en terme de temps à  $10 * 15$  jours soit cinq mois. A l'heure de la rédaction de ce manuscrit, il ne nous est pas possible de réaliser cette analyse, les résultats présentés seront donc moins précis que ceux qu'aurait fournie l'analyse globale.

Le logiciel Tree-Annotator v1.7.4 a été utilisé pour combiner tous les arbres obtenus sous la forme d'un arbre consensus. Au total, 10 000 arbres (300 000 000/30000) ont été produits, 2000 (20%) ont été considérés comme appartenant à la zone de chauffe de la chaîne (Burn-in) et ont été éliminées. La figure B13 représente l'arbre phylogénétique calibré, seules



les datations des nœuds et les probabilités postérieures associées séparant les six groupes sont représentés, afin, de faciliter la lecture de celui-ci.

De façon générale, cet arbre est caractérisé par des probabilités postérieures élevées au niveau des branches externes définissant six groupes taxonomiques (Figure B13). Les relations phylogénétiques entre les groupes sont floues (Figure B13), à l'exception des *Quercus s.s* et des *Protobalanus* qui se trouvent de nouveau très fortement associés (Figures B10, B11 et B13). En l'absence de calibration (Figure B11), la phylogénie issue d'une analyse Bayésienne et d'une recherche de l'arbre le plus vraisemblable apparaît différente de celle obtenue après incorporation de données fossiles (Figure B13). En effet, les *Cyclobalanopsis* se retrouvent soit en groupe frère des *Cerris* (Figure B11) soit seuls, et les *Cerris* se retrouvent en groupe frère des *Ilex* (Figure B13). L'estimation de la longueur des branches d'un arbre phylogénétique ainsi que la topologie de celui-ci sont liés dans les inférences phylogénétiques (Ho and Phillips, 2009). La longueur de branche est fonction d'un taux d'évolution sur un certain temps, ces deux informations sont mélangées. Les données fossiles permettent d'estimer ce temps, et la longueur des branches d'un arbre calibré indique un taux d'évolution qui n'est plus confondu avec le temps, celui-ci se retrouvant en abscisse (Figure B13). L'utilisation de cette nouvelle manière de calculer la longueur de branche influence donc la topologie de l'arbre lorsque les données moléculaires ne sont pas capables de trancher en faveur d'une solution unique (Ho and Phillips, 2009; Sauquet *et al.*, 2012). Cette remarque vient corroborer l'hypothèse du manque de résolution au niveau des nœuds profonds, caractérisée par un faible soutien des branches profondes, dû à un manque d'information du jeu de données moléculaires.

Les données fossiles et celles dérivant de l'analyse bayésienne par BEAST ne sont pas du même ordre de grandeur (Tableau B9). En effet, les données fossiles sont toutes bien plus anciennes que celles estimées (Tableau B9). Si la calibration est de bonne qualité, l'analyse ira dans le même sens que les *a priori* (valeurs indiquées au début de l'analyse) qui prendront des valeurs proches des *à posteriori* (valeurs retournées une fois l'analyse Bayésienne effectuée), dans le cas contraire, il faudra se questionner sur la qualité des données de calibration (Sanders and Lee, 2007). A l'avenir, il faudra tester tous les *a priori* utilisés pour calibrer la phylogénie un par un pour voir si les *a posteriori* correspondent, et éliminer les calibrations introduisant des problèmes avant de lancer une analyse plus complète (Sanders and Lee, 2007). De plus, le fait que la convergence n'ait pas été atteinte dans cette analyse peut être un facteur aggravant de cette incongruence.

TABLEAU B8 : TABLEAU RESUMANT TOUTES LES VALEURS DES ESS (EFFECTIVE SAMPLE SIZE) DE CHACUN DES PARAMETRES MODELISES LORS DE L'ANALYSE PAR LE LOGICIEL BEAST. LES VALEURS EN ROUGES SONT INFERIEURES A 100, CELLES EN ORANGE SONT INFERIEURES A 200.

Statistique	ESS	Statistique	ESS	Statistique	ESS	Statistique	ESS
posterior	1289,97	gene2.frequencies4	4195,77	gene6.frequencies1	4728,577	gene2.coefficientOfVariation	426,507
prior	98,199	gene2.alpha	3220,4	gene6.frequencies2	5721,692	gene2.covariance	9001
likelihood	148,022	gene3.kappa	4046,18	gene6.frequencies3	4822,973	gene3.meanRate	202,152
treeModel.rootHeight	7882,45	gene3.frequencies1	538,381	gene6.frequencies4	4327,042	gene3.coefficientOfVariation	86,313
tmrca(all)	7882,45	gene3.frequencies2	5192,06	gene6.plnv	6896,867	gene3.covariance	3135,561
tmrca(blancs)	52,824	gene3.frequencies3	376,368	gene1.uclid.mean	3001,82	gene4.meanRate	452,735
tmrca(cerris)	86,162	gene3.frequencies4	4923,58	gene1.uclid.stdev	185,216	gene4.coefficientOfVariation	1032,053
tmrca(cyclo)	21,028	gene3.alpha	4542,48	gene2.uclid.mean	1776,885	gene4.covariance	7633,548
tmrca(ilexs)	13,341	gene4.kappa	4266,06	gene2.uclid.stdev	148,098	gene5.meanRate	4482,602
tmrca(proto)	752,055	gene4.frequencies1	4979,47	gene3.uclid.mean	112,437	gene5.coefficientOfVariation	1233,145
tmrca(rouges)	11,458	gene4.frequencies2	5376,73	gene3.uclid.stdev	55,015	gene5.covariance	8266,87
yule.birthRate	1775,83	gene4.frequencies3	5085,83	gene4.uclid.mean	3359,388	gene6.meanRate	3157,875

gene1.kappa	8044,73	gene4.frequencies4	4789,71	gene4.uclid.stdev	2587,262	gene6.coefficientOfVariation	603,213
gene1.frequencies1	5029,02	gene4.alpha	3094,14	gene5.uclid.mean	1835,144	gene6.covariance	8477,731
gene1.frequencies2	4807,11	gene5.kappa	7677,66	gene5.uclid.stdev	975,829	gene1.treeLikelihood	101,184
gene1.frequencies3	5218,62	gene5.frequencies1	4424,56	gene6.uclid.mean	2895,471	gene2.treeLikelihood	56,336
gene1.frequencies4	4787,87	gene5.frequencies2	4572,74	gene6.uclid.stdev	693,225	gene3.treeLikelihood	9,683
gene2.kappa	3211,77	gene5.frequencies3	3514,76	gene1.meanRate	3869,826	gene4.treeLikelihood	216,384
gene2.frequencies1	4462,43	gene5.frequencies4	4644,51	gene1.coefficientOfVariation	171,581	gene5.treeLikelihood	99,549
gene2.frequencies2	4576,81	gene5.alpha	4453,79	gene1.covariance	7623,372	gene6.treeLikelihood	610,134
gene2.frequencies3	4897,87	gene6.kappa	5087,16	gene2.meanRate	1839,393	speciation	129,9

---

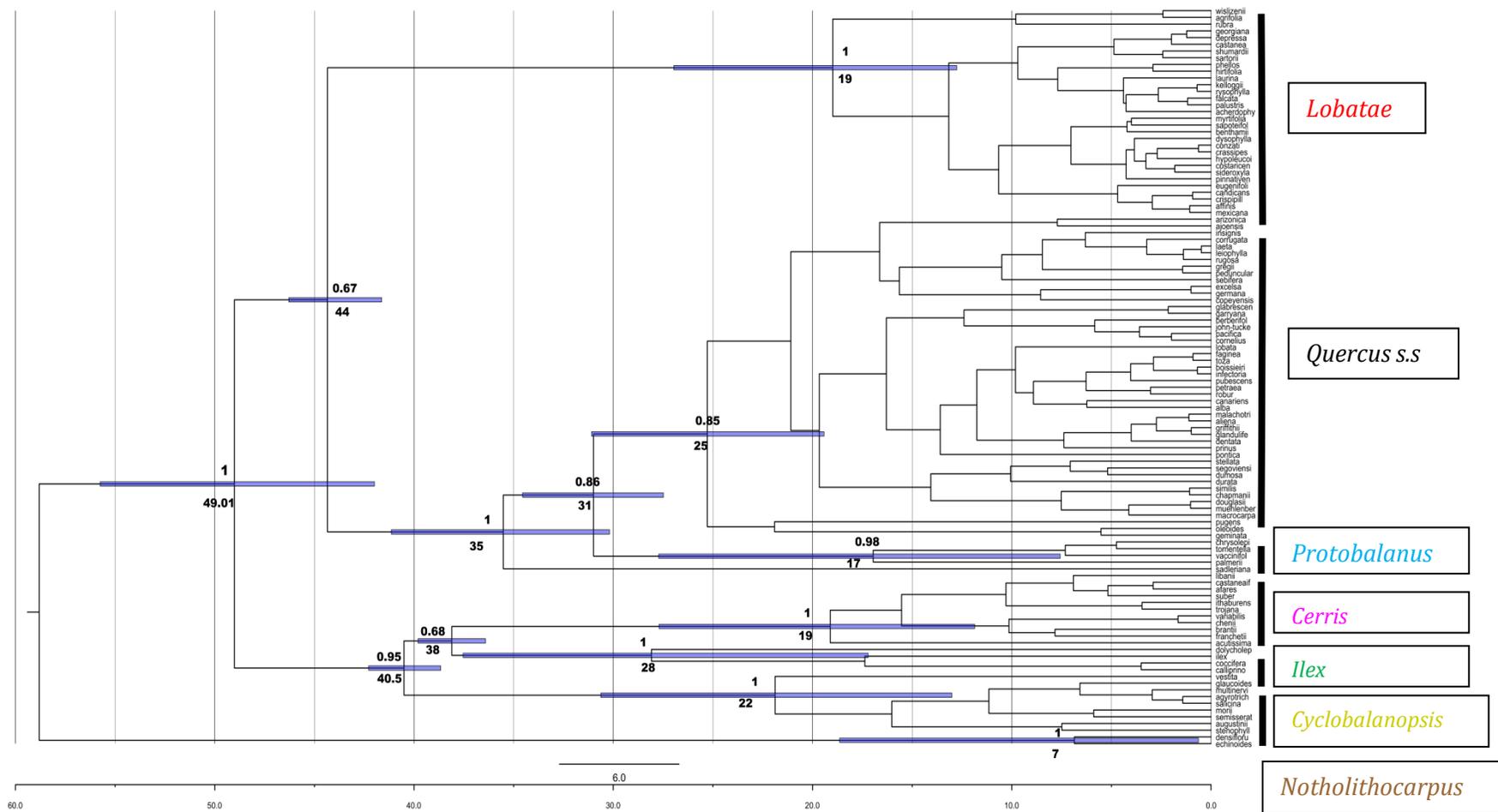


FIGURE B13 : PHYLOGENIE CALIBREE BASEE SUR SIX GENES NUCLEAIRES. LES DATATIONS DES PRINCIPAUX NOEUDS SONT INDIQUEES EN MILLIONS D'ANNEES (AVEC LES BARRES D'ERREURS) AINSI QUE LES PROBABILITES POSTERIEURES ASSOCIEES (RESULTATS OBTENUS PAR BEAST)

TABLEAU B9 : COMPARAISON ENTRE LES DATATIONS FOSSILES UTILISEES POUR LA CALIBRATION DE LA PHYLOGENIE MOLECULAIRE BASEE SUR SIX SEQUENCES NUCLEAIRES ET LES VALEURS RETOURNEES APRES ANALYSE. LES DATES SONT INDIQUEES EN MILLIONS D'ANNEES.

	<b>Calibration fossile</b>	<b>Estimation (Beast)</b>
<b>Lobatae</b>	<b>42-46</b>	<b>12-27</b>
<b>Quercus s.s.</b>	<b>40.4-55.8</b>	<b>18-32</b>
<b>Protobalanus</b>	<b>28.4-33.9</b>	<b>7-27</b>
<b>Cerris</b>	<b>37.2-40.4</b>	<b>12-28</b>
<b>Ilex</b>	<b>37.2-40.4</b>	<b>17-38</b>
<b>Cyclobalanopsis</b>	<b>33.9-55.8</b>	<b>13-31</b>



### B3.7 DEROULEMENT DE CETTE ETUDE

---

Les cartes de distribution des espèces ont été produites par Paul Fromage, actuellement en contrat à l'OAFS (Observatoire Aquitain de la Faune Sauvage) de l'Université Bordeaux 1.

Un soutien logistique important concernant l'utilisation des différents logiciels d'inférences phylogénétiques m'a été apporté par différents intervenants : Alain Franc, Jean-Marc Frigério et Thierry Labbé pour l'utilisation du cluster. De nombreux problèmes informatiques m'ont été évités grâce à l'aide de Loic Kerdraon et à la migration de Windows vers Linux...

J'ai réalisé le travail de laboratoire (séquençage des six gènes nucléaires sur plus de 110 individus) au sein de la Plateforme Génome Transcriptome. La rédaction de l'article inclus dans cette partie a débuté en août 2012 avec la participation très active de mon directeur de thèse (Antoine Kremer). Ces résultats ont été présentés en octobre 2012, lors du congrès international IUFRO. J'ai échangé quelques semaines après avec l'orateur invité Guido Grimm (Chercheur au Muséum d'Histoire Naturelle Suédois) au sujet de mes analyses. Il a attiré notre attention sur le fait que l'espèce servant de racine (*Castanea mollissima*) à nos analyses semblait très éloignée des autres espèces et pouvait induire un phénomène d'attraction des longues branches. C'est à ce moment là qu'il a été décidé de tester une racine alternative (*Notholithocarpus densiflorus*) sur ces six gènes. Le travail de laboratoire a été sous-traité à la Plateforme Génome Transcriptome. Ce travail fait partie intégrante de l'article, et les nombreuses remarques de Guido Grimm à ce sujet nous ont conduits à l'inclure dans la liste des co-auteurs de l'article.

Un soutien important lors des analyses Bayésiennes (MrBayes et BEAST) m'a été apporté lors de mon séjour de quelques jours fin décembre 2012 avec Emmanuelle Jousset (chercheur au CBGP de Montpellier). Plusieurs tests ont été nécessaires pour optimiser au mieux l'obtention d'une phylogénie calibrée, mais comme indiqué plus haut, le temps pour que les données convergent est très long et ne permettra pas de présenter quelque chose de robuste dans les mois à venir.



### B3.8 REFERENCES

---

- Axelrod, D. I. (1983). "Biogeography of Oaks in the Arcto-Tertiary Province." Annals of the Missouri Botanical Garden **70**(4): 629-657.
- Bellarosa, R., *et al.* (2005). "Utility of ITS sequence data for phylogenetic reconstruction of Italian *Quercus* spp." Molecular Phylogenetics and Evolution **34**(2): 355-370.
- Bergsten, J. (2005). "A review of long-branch attraction." Cladistics **21**(2): 163-193.
- Boavida, L. C., *et al.* (2001). "Sexual reproduction in the cork oak (*Quercus sober* L). - II. Crossing intra- and interspecific barriers." Sexual Plant Reproduction **14**(3): 143-152.
- Burgarella, C., *et al.* (2009). "Detection of hybrids in nature: application to oaks (*Quercus suber* and *Q-ilex*)." Heredity **102**(5): 442-452.
- Camus, A. (1936-1954). Les chênes, Monographie du genre Quercus et Monographie du genre Lithocarpus.
- Cottam, W. P., *et al.* (1982). Oak hybridization at the University of Utah. State Arboretum of Utah, Salt Lake City, Utah, USA.
- De la Torre Barcena, J. E. (2009). Seed plant phylogenomics: Phylogeny, character support and structure-function inferences from a multigene analysis. New York, New York University. **Ph.D:** 170.
- Denk, T. and G. W. Grimm (2009). "Significance of Pollen Characteristics for Infrageneric Classification and Phylogeny in *Quercus* (Fagaceae)." International Journal of Plant Sciences **170**(7): 926-940.
- Denk, T. and G. W. Grimm (2010). "The oaks of western Eurasia: Traditional classifications and evidence from two nuclear markers." Taxon **59**(2): 351-366.
- Denk, T., *et al.* (2012). "Fagaceae from the early Oligocene of Central Europe: Persisting new world and emerging old world biogeographic links." Review of Palaeobotany and Palynology **169**(0): 7-20.
- Drummond, A. J., *et al.* (2006). "Relaxed phylogenetics and dating with confidence." Plos Biology **4**(5): 699-710.
- Drummond, A. J., *et al.* (2012). "Bayesian Phylogenetics with BEAUti and the BEAST 1.7." Molecular Biology and Evolution **29**(8): 1969-1973.



- Gouy, M., *et al.* (2010). "SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building." Molecular Biology and Evolution **27**(2): 221-224.
- Guindon, S. and O. Gascuel (2003). "A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood." Systematic Biology **52**(5): 696-704.
- Hall, T. A. (1999). "BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT." Nucleic Acids Symposium Series **41**: 95-98.
- Ho, S. Y. W. and M. J. Phillips (2009). "Accounting for Calibration Uncertainty in Phylogenetic Estimation of Evolutionary Divergence Times." Systematic Biology **58**(3): 367-380.
- Huelsenbeck, J. P. and F. Ronquist (2001). "MRBAYES: Bayesian inference of phylogenetic trees." Bioinformatics **17**(8): 754-755.
- Hughes, C. E., *et al.* (2006). "From famine to feast? Selecting nuclear DNA sequence loci for plant species-level phylogeny reconstruction." Philosophical Transactions of the Royal Society B-Biological Sciences **361**(1465): 211-225.
- Kremer, A., *et al.* (2012). "Genomics of Fagaceae." Tree Genetics & Genomes **8**(3): 583-610.
- Kremer, A. and R. Petit (1993). "Gene diversity in natural populations of oak species." Ann. For. Sci. **50**(Supplement): 186s-202s.
- Lumaret, R. and R. Jabbour-Zahab (2009). "Ancient and current gene flow between two distantly related Mediterranean oak species, *Quercus suber* and *Q-ilex*." Annals of Botany **104**(4): 725-736.
- Manos, P. S., *et al.* (2008). "Phylogenetic Relationships and Taxonomic Status Of the Paleoendemic Fagaceae Of Western North America: Recognition Of A New Genus, *Notholithocarpus*." Madroño **55**(3): 181-190.
- Manos, P S., *et al.* (1999). "Phylogeny, Biogeography, and Processes of Molecular Differentiation in *Quercus* Subgenus *Quercus* (Fagaceae)." Molecular Phylogenetics and Evolution **12**(3): 333-349.
- Manos, P. S. and A. M. Stanford (2001). "The Historical Biogeography of Fagaceae: Tracking the Tertiary History of Temperate and Subtropical Forests of the Northern Hemisphere." International Journal of Plant Sciences **162**(S6): S77-S93.



- Manos, P. S., *et al.* (2001). "Systematics of Fagaceae: Phylogenetic Tests of Reproductive Trait Evolution." International Journal of Plant Sciences **162**(6): 1361-1379.
- Menitsky, J. L. (1984). Duby Asii. Leningrad: Nauka. .
- Menitsky, Y. L. (2005). Oaks of Asia, Science Publishers.
- Mir, C., *et al.* (2006). "Endemic North African *Quercus afares* Pomel originates from hybridisation between two genetically very distant oak species (*Q. suber* L. and *Q. canariensis* Willd.): evidence from nuclear and cytoplasmic markers." Heredity (Edinb) **96**(2): 175-184.
- Moran, E. V., *et al.* (2012). "Genetic evidence for hybridization in red oaks (*Quercus* sect. *Lobatae*, Fagaceae)." American Journal of Botany **99**(1): 92-100.
- Nixon, K. (1993). "Infrageneric classification of *Quercus* (Fagaceae) and typification of sectional names." Ann. For. Sci. **50**(Supplement): 25s-34s.
- Oh, S. H. and P. S. Manos (2008). "Molecular phylogenetics and cupule evolution in Fagaceae as inferred from nuclear CRABS CLAW sequences." Taxon **57**(2): 434-451.
- Ørsted, A. S. (1867). Bidrag til Egeslægtens Systematik.
- Posada, D. and K. A. Crandall (1998). "MODELTEST: testing the model of DNA substitution." Bioinformatics **14**(9): 817-818.
- Rambaut, A. and A. Drummond (2007). Tracer v1.4, Available from <http://beast.bio.ed.ac.uk/Tracer>.
- Rehder, A. (1940). Manual of cultivated trees and shrubs hardy in North America: exclusive of the subtropical and warmer temperate regions, The Macmillan Company.
- Rokas, A. and S. B. Carroll (2005). "More genes or more taxa? The relative contribution of gene number and taxon number to phylogenetic accuracy." Molecular Biology and Evolution **22**(5): 1337-1344.
- Rokas, A., *et al.* (2005). "Animal evolution and the molecular signature of radiations compressed in time." Science **310**(5756): 1933-1938.
- Rushton, B. (1993). "Natural hybridization within the genus *Quercus* L." Ann. For. Sci. **50**(Supplement): 73s-90s.
- Samuel, R., *et al.* (1998). "ITS sequences from nuclear rDNA suggest unexpected phylogenetic relationships between Euro-Mediterranean, East Asiatic and North



- American taxa of *Quercus* (Fagaceae)." Plant Systematics and Evolution **211**(1-2): 129-139.
- Sanders, K. L. and M. S. Y. Lee (2007). "Evaluating molecular clock calibrations using Bayesian analyses with soft and hard bounds." Biology Letters **3**(3): 275-279.
- Sauquet, H., *et al.* (2012). "Testing the Impact of Calibration on Molecular Divergence Times Using a Fossil-Rich Group: The Case of *Nothofagus* (Fagales)." Systematic Biology **61**(2): 289-313.
- Schnitzler, J. P., *et al.* (2004). "Hybridization of European oaks (*Quercus ilex* x *Q-robur*) results in a mixed isoprenoid emitter type." Plant Cell and Environment **27**(5): 585-593.
- Schwarz, O. (1936-39). Monographie der Eichen Europas und des Mittelmeergebietes. Berlin.
- Schwarz, O. (1993). *Quercus* L. Flora Europaea. T. G. Tutin, V. H. Heywood, N. A. Burges *et al.* Cambridge, Cambridge University Press. **1 : Lycopodiaceae to Platanaceae**.
- Scornavacca, C., *et al.* (2008). "PhySIC\_IST: cleaning source trees to infer more informative supertrees." Bmc Bioinformatics **9**.
- Stamatakis, A., *et al.* (2008). "A Rapid Bootstrap Algorithm for the RAxML Web Servers." Systematic Biology **57**(5): 758-771.
- Tamura, K., *et al.* (2007). "MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0." Mol Biol Evol **24**(8): 1596-1599.
- Ueno, S., *et al.* (2010). "Bioinformatic analysis of ESTs collected by Sanger and pyrosequencing methods for a keystone forest tree species: oak." Bmc Genomics **11**: 650.
- Valencia, S. A. (2004). "Diversidad del genero *Quercus* (Fagaceae) en Mexico. ." Bol. Soc. Bot. Mex. **75**: 33-54.
- Whitfield, J. B. and K. M. Kjer (2008). "Ancient rapid radiations of insects: Challenges for phylogenetic analysis." Annual Review of Entomology **53**: 449-472.
- Xu, L. (2004). Diversité de l'ADN chloroplastique et relations phylogénétiques au sein des Fagacées et du genre *Quercus*. Thèse de l'Université Henri Poincaré, Nancy, France.
- Zhang, D. X. and G. M. Hewitt (1996). "Nuclear integrations: Challenges for mitochondrial DNA markers." Trends in Ecology & Evolution **11**(6): 247-251.



## B4 ETUDE DU GENOME CHLOROPLASTIQUE

---



## B4.1 INTRODUCTION

---

Comme nous l'avons vu dans la partie précédente, l'utilisation de six nouvelles séquences nucléaires a permis de clarifier les relations entre certains groupes taxonomiques, mais s'est heurtée au même problème que celui rencontré avec d'autres marqueurs (ITS et CRC notamment) : un manque de résolution au niveau des nœuds profonds. Le manque de soutien aux branches profondes permet de soupçonner cette irrésolution. L'hypothèse d'un signal phylogénétique faible, dont les causes peuvent être soit techniques, soit biologiques (spéciation rapide), est donc toujours présente. L'utilisation des nouvelles technologies de séquençage (Next Generation Sequencing Technologies) permettant d'explorer une partie importante du génome est une voie pour résoudre le manque de résolution aux nœuds internes chez les plantes (Hughes *et al.*, 2006). Par contre si l'hypothèse de radiation rapide est la cause de la faible résolution, elle a dû marquer de son empreinte l'ensemble du génome, non seulement nucléaire mais aussi cytoplasmique.

Cette question est à la base de la présente étude : à travers le séquençage total du génome chloroplastique (plus de 160 000 pb) d'une soixantaine d'espèces, plusieurs types de reconstructions phylogénétiques vont être utilisées : la méthode SuperMatrice (sur le génome total et sur une sous partie de celui-ci) et la méthode SuperArbre (sur plusieurs dizaines de gènes). La comparaison de ces deux méthodes devrait pouvoir indiquer si oui ou non une radiation rapide peut expliquer l'irrésolution retrouvée au niveau des nœuds profonds dans le génome nucléaire et chloroplastique.

Le génome chloroplastique a été l'un des premiers à être utilisé en génétique des populations, à cause notamment du très grand nombre de copies présentes dans le génome du chloroplaste (Bendich, 1987). Actuellement, des séquences du génome chloroplastique entier de plus d'une centaine d'espèces d'angiospermes sont disponibles dans les banques de données (Kremer *et al.*, 2012). Le génome chloroplastique est constitué par une centaine de gènes pouvant être séparés en trois types de régions génomiques (Hughes *et al.*, 2006) : des gènes codant pour des protéines, des introns et des espaceurs intergéniques. Les deux derniers ne codent pas pour des protéines et sont donc des régions non codantes. Ce génome est en fait une molécule d'ADN circulaire double brin, séparé en trois régions. Chez le chêne (*Quercus spp*), ces trois régions ont une taille variant de 90 000 pb pour la grande région simple copie (Large-Single Copy, LSC), à 19 000 pb pour la petite région simple copie (Small-Single Copy, SSC), et de 25 000 pb pour les deux régions répétées inversées (Inverted Repeat, IR) (Kremer *et al.*, 2012).



Les études précédentes suggèrent que les régions répétées inversées accumulent moins de mutations que les régions simples copies (Curtis and Clegg, 1984; Wolfe *et al.*, 1987; Gaut, 1998; Hughes *et al.*, 2006) ; ces dernières accumulant 2.3 fois plus de substitutions que les régions répétées (Perry and Wolfe, 2002). L'hérédité maternelle du génome chloroplastique a été montrée chez les chênes par l'analyse de la ségrégation des variants chloroplastiques dans des familles issues de croisements contrôlés (Dumoulin *et al.*, 1995). Enfin, au cours de la transmission maternelle, la molécule d'ADN chloroplastique est reproduite de manière clonale et n'est pas sujette à la recombinaison. Ces propriétés (transmission maternelle, reproduction clonale, absence de recombinaison, taux d'évolution donnant un signal adéquat) font de l'ADN chloroplastique un marqueur très intéressant en phylogénie et surtout en phylogéographie (étude des principes et processus qui gouvernent la distribution des lignées généalogiques, spécialement celle de niveau intraspécifique). En effet comme les espèces végétales ne peuvent migrer que par les graines (porteur de l'ADN chloroplastique du parent), l'empreinte de la migration peut se retrouver dans la distribution géographique actuelle des variants de l'ADN chloroplastique. Ces prédictions sont vérifiées par la comparaison de la variation du polymorphisme de l'ADN chloroplastique chez les chênes blancs européens d'une part et les reconstructions des voies de migrations postglaciaires faites sur la base de données historiques (pollens fossiles) d'autre part (Petit *et al.*, 2002). Ces résultats obtenus sur l'ensemble de l'Europe et sur une échelle de temps couvrant l'holocène, ont démontré que l'ADN chloroplastique constituait un outil puissant de résolution phylogéographique au niveau de la microévolution. Il apparaissait donc très judicieux de vérifier si l'ADN chloroplastique total pouvait conduire à la même résolution au niveau de la biogéographie historique du genre (analyse des relations entre la structure et l'histoire des espèces d'une part, et l'histoire géologique de la surface du globe, d'autre part). De manière plus explicite, est-ce que la phylogénie de l'ADN chloroplastique est structurée selon la géographie ? Telle est la question à laquelle ce chapitre tente de répondre également. Des résultats antérieurs obtenus sur un fragment de l'ADN chloroplastique de 1700 pb suggéraient déjà une forte structure géographique (Xu, 2004). L'analyse sur le génome chloroplastique entier devrait permettre de conforter ces premières observations. La mise en évidence d'une forte structuration géographique permettrait de faciliter l'interprétation de l'histoire biogéographique de l'espèce en suggérant des voies de migration au cours du tertiaire : c'est là le véritable intérêt de l'apport de l'ADN chloroplastique.



Pour résumer, l'objectif que souhaite atteindre cette étude est donc triple :

- 1) Vérifier si l'accès à la séquence du génome chloroplastique entier permet d'améliorer la résolution au niveau des nœuds profonds,
- 2) Et, si tel est le cas, vérifier si la prospection du génome chloroplastique peut être limitée à une fraction plus réduite. Nous essaierons notamment de voir si les séquences d'un nombre réduit de gènes du génome chloroplastique peuvent fournir un signal phylogénétique au moins équivalent à celui issu du génome entier,
- 3) Vérifier si la phylogénie de l'ADN chloroplastique suit une organisation géographique (structure phylogéographique).

TABLEAU B10 : ORIGINE GEOGRAPHIQUE DES ESPECES ET TAILLE RELATIVE DU FRAGMENT AMPLIFIE.

<b>Nom Espèce</b>	<b>Classification</b>	<b>Taille/164928</b>	<b>Origine géographique</b>	<b>Latitude</b>	<b>Longitude</b>
<i>Castanea sativa</i>	<i>Castanea</i>	0,975207363	Italie	40°04'N	11°37'E
<i>Quercus afares</i>	<i>Cerris</i>	0,960552484	Algérie	36°46'N	3°40'E
<i>Quercus acutissima</i>	<i>Cerris</i>	0,974728366	Chine	32°04'N	118°05'E
<i>Quercus chenii</i>	<i>Cerris</i>	0,973079162	Chine	32°04'N	118°50'E
<i>Quercus crenata</i>	<i>Cerris</i>	0,978129851	France	43°08'N	10°51'E
<i>Quercus ithaburensis</i>	<i>Cerris</i>	0,977875194	Italie	40°37'N	17°10'E
<i>Quercus suber</i>	<i>Cerris</i>	0,977772119	France	44°38'N	1°10'W
<i>Quercus trojana</i>	<i>Cerris</i>	0,977499272	Italie	40°37'N	17°10'E
<i>Quercus variabilis</i>	<i>Cerris</i>	0,975595411	Chine	32°04'N	118°50'E
<i>Quercus argyrotricha</i>	<i>Cyclobalanopsis</i>	0,697813591	Chine		
<i>Quercus morii</i>	<i>Cyclobalanopsis</i>	0,963153619	Japon		
<i>Quercus multinervis</i>	<i>Cyclobalanopsis</i>	0,706605307	Chine		
<i>Quercus salicina</i>	<i>Cyclobalanopsis</i>	0,655546663	Japon	34°45' N	133°45'E
<i>Quercus stenophylloides</i>	<i>Cyclobalanopsis</i>	0,686487437	Japon		
<i>Quercus alnifolia</i> Poech	<i>Ilex</i>	0,794346624	Chypre		
<i>Quercus calliprinos</i>	<i>Ilex</i>	0,977747866	Israël	32° 40'N	35°05'E
<i>Quercus franchetii</i>	<i>Ilex</i>	0,824929666	Chine	25°05'N	102°26'E
<i>Quercus ilex</i>	<i>Ilex</i>	0,977996459	France	44°38'N	1°10'W
<i>Quercus semecarpifolia</i>	<i>Ilex</i>	0,754347352	Chine	25°05'N	102°26'E
<i>Quercus crispipilis</i>	<i>Lobatae</i>	0,755081005	Mexique	16°47'N	92°5'W
<i>Quercus hypoleucooides</i>	<i>Lobatae</i>	0,712620052	USA (Nouveau Mexique)	35°19'N	106°50'W
<i>Quercus kelloggii</i>	<i>Lobatae</i>	0,975777309	USA (Californie)	37°36'N	120°02'W
<i>Quercus mexicana</i>	<i>Lobatae</i>	0,974534342	Mexique	20°07'N	98°18'W
<i>Quercus nigra</i>	<i>Lobatae</i>	0,975116414	US(Arkansas)	33°05'N	93°30'W
<i>Quercus phellos</i>	<i>Lobatae</i>	0,973018529	USA (Louisiane)	32°30'N	92°40'W
<i>Quercus rysophylla</i>	<i>Lobatae</i>	0,655613359	Mexique	25°36'55"	100°21'54"
<i>Quercus sartorii</i>	<i>Lobatae</i>	0,683601329	Mexique	20° 38'N	98° 36'W
<i>Quercus shumardii</i>	<i>Lobatae</i>	0,967773816	USA (Arkansas)	33°05'N	93°30'W

<i>Quercus wislizeni</i>	<i>Lobatae</i>	0,820776339	USA (Californie)	33°42'N	117°25'W
<i>Quercus benthamii</i>	<i>Lobatae</i>	0,97122987	Mexique	17°05'N	93°2'W
<i>Quercus chrysolepis</i>	<i>Protobalanus</i>	0,70139091	USA (Californie)	37°10'N	121°25'W
<i>Quercus palmeri</i>	<i>Protobalanus</i>	0,925706975	USA (Californie)	33°35'41"	116°36'32"
<i>Quercus toza</i>	<i>Quercus s.s.</i>	0,892431849	France	44°44'N	0°45'W
<i>Quercus aliena</i>	<i>Quercus s.s.</i>	0,976905074	Chine	32°04'N	118°50'E
<i>Quercus canariensis</i>	<i>Quercus s.s.</i>	0,97445552	Algérie	36°47'N	5°46'E
<i>Quercus copeyensis</i>	<i>Quercus s.s.</i>	0,710115929	Costa Rica	9°37'N	83°35' W
<i>Quercus dalechampii</i>	<i>Quercus s.s.</i>	0,978057092	Italie	40°21'N	18°10'E
<i>Quercus dentata</i>	<i>Quercus s.s.</i>	0,971666424	Chine	32°04'N	118°50'E
<i>Quercus douglasii</i>	<i>Quercus s.s.</i>	0,772937282	USA (Californie)	38°43'N	120°48'W
<i>Quercus durata</i>	<i>Quercus s.s.</i>	0,800913126	USA (Californie)	38°59'20"	123°05'07"
<i>Quercus fabri</i>	<i>Quercus s.s.</i>	0,970229433	Chine	32°04'N	118°50'E
<i>Quercus faginea</i>	<i>Quercus s.s.</i>	0,928192908	Espagne	42°34'N	0°34'W
<i>Quercus insignis</i>	<i>Quercus s.s.</i>	0,806764164	Mexique	19° 55'N	96°58'W
<i>Quercus laeta</i>	<i>Quercus s.s.</i>	0,923742482	Mexique	17° 25'N	96° 35'W
<i>Quercus lusitanica*</i>	<i>Quercus s.s.</i>	0,97816623	Espagne	42°34'N	0°34'W
<i>Quercus oleoides</i>	<i>Quercus s.s.</i>	0,674803551	Cuba	22°37'N	83°39'W
<i>Quercus pacifica</i>	<i>Quercus s.s.</i>	0,697777212	USA (Ile de Santa cruz)	33°59' N	119°43'W
<i>Quercus petraea</i>	<i>Quercus s.s.</i>	0,978008586	France	47°55'N	1°54'E
<i>Quercus polymorpha</i>	<i>Quercus s.s.</i>	0,670522895	Mexique		
<i>Quercus montana</i>	<i>Quercus s.s.</i>	0,976632227	USA (New York)	43°16'N	77°37'W
<i>Quercus pubescens</i>	<i>Quercus s.s.</i>	0,977875194	France	45°18'N	0°54'W
<i>Quercus robur</i>	<i>Quercus s.s.</i>	0,978087408	France	44°44'N	0°42'W
<i>Quercus rugosa</i>	<i>Quercus s.s.</i>	0,882918607	Mexique	17° 25'N	99° 35'W
<i>Quercus sadleriana</i>	<i>Quercus s.s.</i>	0,833685002	USA (Californie)	41°59'N	123°31'W
<i>Quercus sebifera</i>	<i>Quercus s.s.</i>	0,822534682	Mexique	16°05'N	92°00'W
<i>Quercus segoviensis</i>	<i>Quercus s.s.</i>	0,773955908	Mexique	16°47'N	92°05'W
<i>Quercus virgiliana**</i>	<i>Quercus s.s.</i>	0,977335565	Italie	40°21'N	18°10'E

\**Quercus lusitanica* est appelé *Quercus faginea* dans Govaerts et Frodin (Govaerts and Frodin, 1998).

\*\**Quercus virgiliana* est appelé *Quercus pubescens* dans Govaerts et Frodin (Govaerts and Frodin, 1998).



## B4.2 MATERIEL ET METHODES

---

### B4.2.1 ECHANTILLONNAGE

---

Dans le cadre du projet Européen Evoltree (<http://www.evoltree.eu>), le séquençage total du génome chloroplastique a été mis au point. Ces travaux préliminaires ont porté sur plusieurs espèces au tout début afin d'évaluer la faisabilité de la méthode : *Quercus petraea*, *Q. suber*, *Q. ilex*, *Q. macrolepis*, *Q. trojana*, *Q. calliprinos*, *Q. dalechampii*, *Q. virginiana*, *Q. pubescens*, *Q. lusitanica* and *Q. crenata*). C'est pourquoi, une partie des espèces utilisées dans les deux parties précédentes a été complétée par ces espèces. L'échantillonnage sur le terrain s'est toujours réalisé de la façon suivante : des feuilles (ou bourgeons) de un à sept arbres par espèce et séparés de moins de 500 mètres ont été récoltés et mises immédiatement en présence de silicagel afin de les sécher, puis elles ont été conservées à -80°C en vue de l'extraction de leur ADN.

Au total, le génome chloroplastique entier de 57 espèces de *Quercus* et de *Castanea* a été séquençé (Tableaux B10 et B11, Figures B14 et B15).

Le groupe des *Quercus s.s* est assez bien représentés car il contient trois espèces asiatiques sur un total de 15-25 (Tableau B11), huit espèces d'Europe et d'Afrique du Nord sur un total de 16 (Tableau B11), et 14 espèces américaines sur un total de 104 (Tableau B11).

Pour le groupe des *Lobatae*, 11 espèces sur 82 ont été échantillonnées, deux sur cinq espèces de *Protobalanus* ont été récoltées et 10 espèces du groupe *Cerris* sur un total compris entre 12 et 32 (Tableau B11) ont été récoltées.

Trois sur quatre espèces du groupe *Ilex* d'Europe et d'Afrique du Nord ont été récoltées (Tableau B11), mais seulement deux sur les 26 espèces asiatiques l'ont été ; ceci peut constituer un trou dans notre échantillonnage.

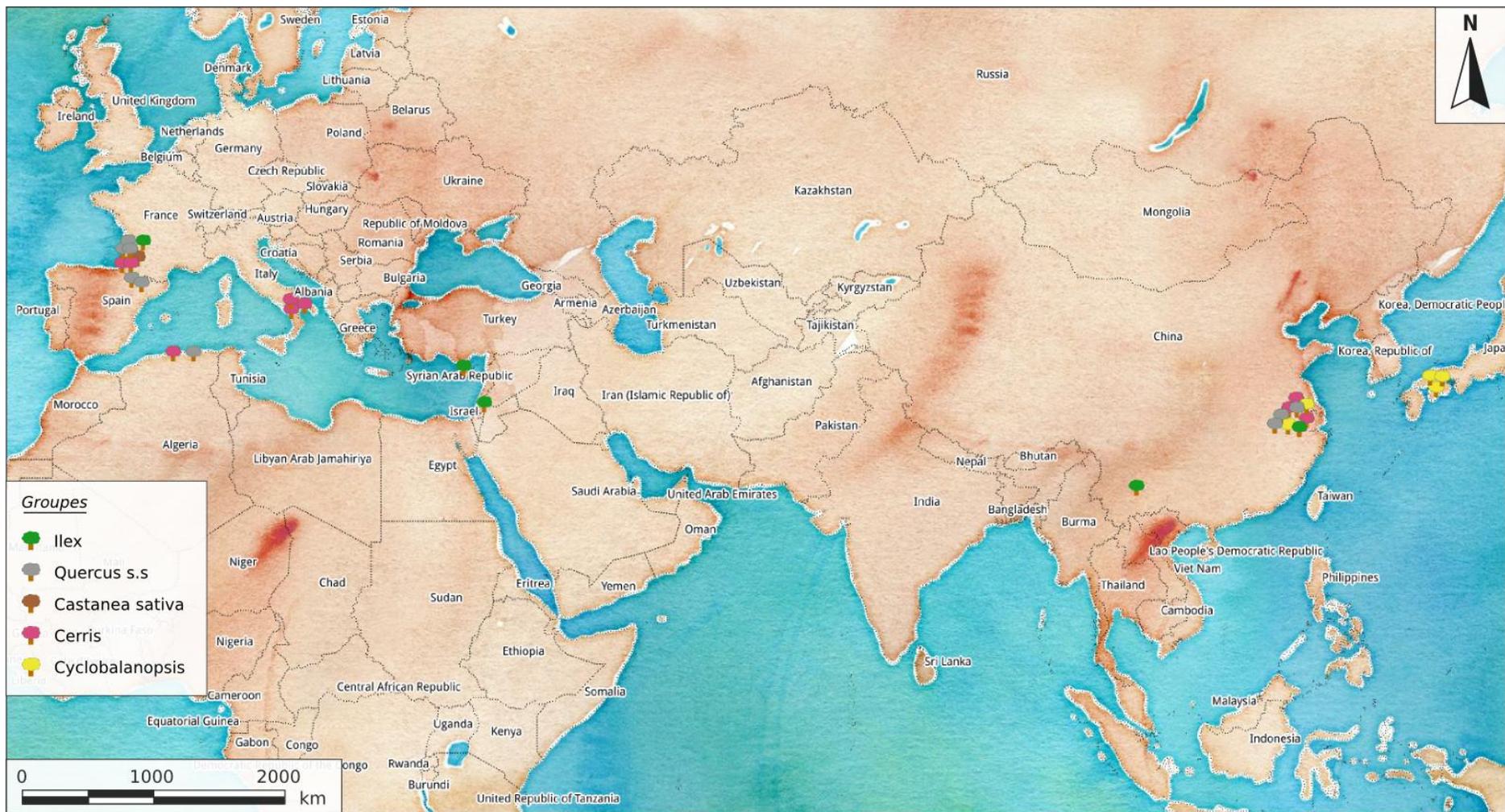


FIGURE B14 : REPARTITION DES ESPECES ECHANTILLONNEES EN EURASIE.

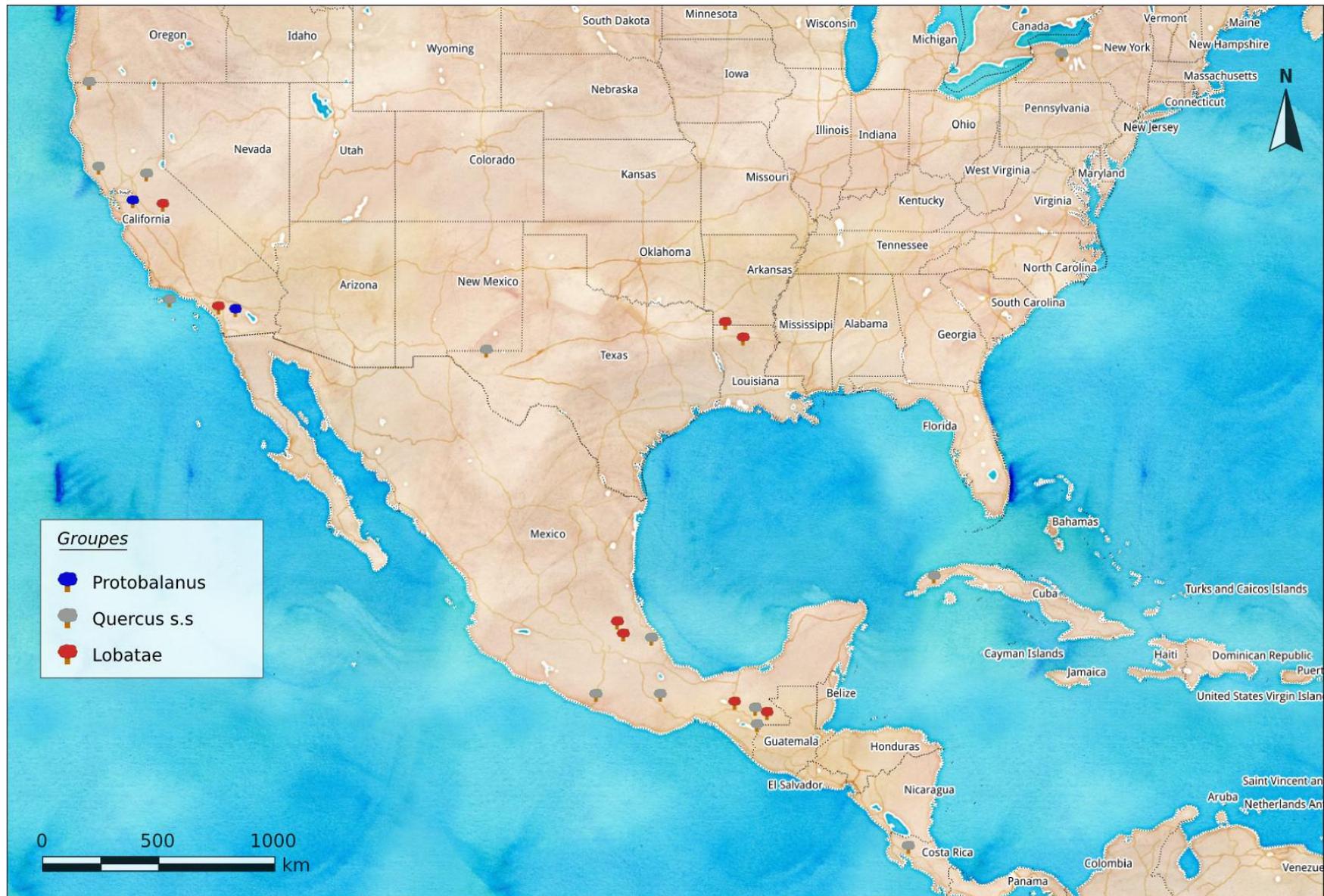


FIGURE B15 : REPARTITION DES ESPECES ECHANTILLONNEES EN AMERIQUE.

TABLEAU B11 : NOMBRE D'ESPECES ANALYSEES ET NOMBRE TOTAL D'ESPECES DU GENRE QUERCUS REPERTORIEES DANS LA LITTERATURE.

Genre	Sous-genre	Groupe	Origine Géographique	Nombre d'espèces	Nombre estimé d'espèces
<i>Castanea</i>			Europe	1	12
<i>Quercus</i>					
	<i>Cyclobalanopsis</i>		Asie	5	87e, 76c
	<i>Quercus s.l.</i>				
		<i>Quercus s.s.</i>	Asie	3	25e, 15c
			Europe et Afrique du Nord	8	16d
			Amérique	14	104a
	<i>Lobatae</i>		Amérique	11	82a
	<i>Protobalanus</i>		Amérique	2	5b
	<i>Cerris</i>		Asie	6	8c
			Europe et Afrique du Nord	4	4d
	<i>Ilex</i>		Asie	2	26c
			Europe et Afrique du Nord	3	4d
				57	

(a) Nixon K.C., 1993b.

(b) Floras (Brach and Song, 2006),( <http://www.efloras.org> ).

(c) Menitsky Y.L. 2005.

(d) Flora europea (<http://rbg-web2.rbge.org.uk/FE/fe.html>).

(e) Camus (1936-1954).

---

## B4.2.2 AMPLIFICATION DE L'ADN CHLOROPLASTIQUE ET SEQUENCAGE

---

### OBTENTION D'AMORCES PERMETTANT L'AMPLIFICATION DU GENOME CHLOROPLASTIQUE TOTAL

---

Le séquençage du génome chloroplastique de *Quercus robur* et de *Castanea sativa* a été réalisé en mettant au point une série d'amorces de PCR qui permettent d'amplifier des régions génomiques se chevauchant partiellement et recouvrant l'ensemble du génome. Tout ce travail a été réalisé à l'Istituto di Genomica Applicata (IGA), dans le cadre d'une coopération avec le CNR (DR G.G. Vendramin).

Concernant la grande région simple copie (LSC), 38 paires d'amorces localisées dans les régions conservées et au préalable décrites par Grivet et al (Grivet *et al.*, 2001) ont été utilisées, en évitant les trous entre les fragments le long de la molécule d'ADN chloroplastique. De nouvelles amorces ont été mises au point pour les régions qui amplifiaient mal ou produisaient des doubles bandes, en suivant la stratégie ci-dessous.

Les amorces des régions SSC et IR ont été dérivées des séquences conservées après alignement avec le logiciel Muscle version 3.7 (Edgar, 2004) chez les génomes chloroplastiques provenant de GenBank de *Jasminum nudiflorum* (DQ673255), *Populus trichocarpa* (EF489041), *Vitis vinifera* (DQ424856), *Eucalyptus globulus* (AY780259), *Arabidopsis thaliana* (AP000423), *Gossypium hirsutum* (DQ345959), *Citrus sinensis* (DQ864733), *Cucumis sativus* (AJ970307), *Morus indica* (DQ226511), *Panax ginseng* (AY582139), *Solanum lycopersicum* (AM087200) and *Nicotiana tabacum* (Z00044).

### REACTIONS D'AMPLIFICATIONS

---

Les réactions d'amplifications ont été faites dans un volume final de 20 µl contenant 10-40 ng d'DNA, 1.5 mM MgCl<sub>2</sub>, 200 µM de chaque dNTP, 0.2 µM de chaque primer et 1 U de GOTaq polymérase (Promega). Les amplifications ont été réalisées dans une machine PCR "PCR System 9700" (Applied Biosystems) en utilisant les conditions d'amplifications suivantes : 3min à 94°C; 35 cycles de 20 s à 95°C, 30 s touchdown 60-50°C pour les 10 premiers cycles et 20-90 s à 72°C; suivis par 10min at 72°C.

Les produits de PCR ont été chargés sur un gel d'agarose 1.5%, colorés au bromure d'éthidium et visualisés sous UV afin de vérifier la bonne amplification. Pour les fragments mesurant plus de 5 000 pb, 1 µl de l'enzyme Herculase II Fusion DNA Polymerase (Agilent) a été utilisée en suivant les consignes du fabricant.

TABLEAU B12 : CARACTERISTIQUES DE GENOME TOTAL ET DES 32 ZONES CHLOROPLASTIQUES EXTRAITES DE L'ALIGNEMENT TOTAL.

Zone	Catégories	Nombre de séquences	Longueur (pb)	Nb de sites Variables*/Taille	Nb de sites Parcimonieux*/Taille	Nb de pattern trouvés (Phyml)	Pourcentage de sites non poly- -morphes
<b>Génome total</b>		<b>57</b>	<b>164928</b>	<b>0.007</b>	<b>0.0027</b>	<b>20460</b>	<b>94</b>
matK	Gène	56	1849	0.012	0.0065	489	90
ndhA	Gène	57	2140	0.01	0.005	252	94
ndhF-rpl32	Zone intergénique	53	1019	0.022	0.0088	276	88
ndhF	Gène	57	2298	0.0074	0.0004	414	91
petA-psbJ	Zone intergénique	57	1065	0.0028	0.0019	259	91
psbE-petL	Zone intergénique	57	1379	0.011	0.0014	349	88
rpl32-trnL	Zone intergénique	56	1206	0.0066	0.0017	322	91
rpoB-trnC	Zone intergénique	56	1289	0.007	0.0023	290	93
rps16-trnQ	Zone intergénique	54	1597	0.0012	0	405	91
trnK	Zone intergénique	57	2589	0.012	0.0043	322	94
trnT-GGU-psbD	Zone intergénique	57	1682	0.012	0.0036	438	91
ycf1	Gène	57	5701	0.0045	0.002	994	93
ycf3	Gène	57	2006	0.012	0.0055	283	94
accD-psaI	Zone intergénique	54	708	0.0085	0	184	93
atpF	Gène	57	1358	0.0007	0	243	94
atpH-atpI	Zone intergénique	57	1205	0.0033	0.0008	337	92

matK	Gène	57	1515	0.018	0.0053	177	95
ndhD	Gène	57	1501	0.018	0.0067	174	95
petB	Gène	57	1491	0.004	0.0027	144	96
psaA-ycf3	Zone intergénique	56	777	0.006	0	200	92
psbZ-trnG	Zone intergénique	53	650	0.003	0.003	219	88
rbcL	Gène	57	1428	0.023	0.012	82	97
rpoB	Gène	57	3222	0.012	0.0043	216	97
rpoC1	Gène	57	2889	0.014	0.0055	238	96
rpoC2	Gène	57	4190	0.02	0.0067	287	96
rps16	Gène	57	1193	0.009	0.0042	194	95
rps4-trnT	Zone intergénique	56	351	0	0	93	87
trnC-GCA-petN	Zone intergénique	57	984	0.0091	0.004	164	95
trnE-UUC-trnT	Zone intergénique	57	1026	0.025	0.015	195	93
trnH-GUG-psbA	Zone intergénique	55	458	0.032	0.015	138	86
trnS-GCU-trnG	Zone intergénique	56	1038	0	0	372	86
ycf3-trnS	Zone intergénique	57	941	0.0096	0.0042	192	93
<b>Total</b>			<b>52745</b>	<b>0.011</b>	<b>0.0041</b>		

\* Le calcul du nombre de sites variables et parcimonieux est effectué sur les données présentes chez toutes les espèces seulement (les données manquantes chez une espèce excluent tous ces sites chez les autres espèces).



Les produits de PCR ont été découpés par nébulisation, puis classés suivant leur taille sur un gel d'électrophorèse. Les produits d'une taille comprise entre 0.5 et 1.5 Kb ont été clonés (Zero Background™/Kan Cloning Kit, Invitrogen) et transformés par des bactéries *E. coli* (Invitrogen) compétentes au sein d'un plasmide. Ces cellules ont été sélectionnées sur un gel de LB agar contenant 40 µg/mL de kanamycine. Les réactions de séquençage ont été effectuées dans les deux sens sur un ABI 3130 xL (Applied Biosystems, Foster City, CA). Les séquences ont été assemblées par CodonCode Aligner. Pour vérifier l'assemblage et l'annotation, les jonctions entre les LSC et les IR, et les SSC et les IR ont été confirmées par séquençage sanger. L'assemblage était considéré comme correct lorsque les séquences de bonnes qualités (scores de KB Basecaller (Applied Biosystems)  $\geq 20$ ) étaient présentes plus de 4 X.

## RESEQUENCAGE

---

L'ADN génomique total a été extrait en utilisant le kit commercial Invisorb DNA plant HTS en plaque de 96 individus en utilisant les recommandations du fabricant. Par la suite, plusieurs amplicons par individu extrait ont été réalisés par PCR dans les mêmes conditions que précédemment. Après une évaluation de l'amplification par visualisation sur gel d'agarose 1%, tous les amplicons ayant amplifié ont été rassemblés et concentrés de façon équimolaire afin d'avoir une quantité finale d'ADN comprise entre 200 et 300 ng, une quantité variable d'ADN total a été ajoutée afin d'amplifier les régions qui n'ont pas amplifiées lors de la PCR. L'ADN chloroplastique ainsi traité a été amplifié sur des lames du séquenceur Illumina (Illumina 1G/Solexa, Illumina Inc., San Diego, CA, USA). Le séquençage a été réalisé à l'Istituto di Genomica Applicata (IGA), dans le cadre d'une coopération avec le CNR (DR G.G. Vendramin). La préparation de tous les produits a été réalisée en suivant le protocole du kit.

TABLEAU B13 : PARAMETRES D'ENTREE DU LOGICIEL PHYSIC-IST (VALEUR DE BOOTSTRAPS ET « SOURCE TREE CORRECTION ») ET PARAMETRES DE SORTIE (NOMBRE D'INDIVIDUS ELIMINES DE L'ANALYSE / VALEUR DE « CLADISTIC INFORMATION CONTENT »). LES VALEURS EN GRAS INDIQUENT LA COMBINAISON AYANT CONDUIT A LA MEILLEURE RESOLUTION.

% Bootstrap \ STC	STC				
	0.1	0.2	0.4	0.6	0.8
40 %	2/0.858	2/0.858	2/0.858	<b>0/0.9240</b>	0/0.914
50 %	0/0.898	0/0.898	0/0.898	<b>0/0.906</b>	0/0.898
60 %	0/0.898	0/0.898	0/0.898	0/0.898	<b>0/0.90219</b>
70 %	0/0.8477	0/0.8477	<b>0/0.8998</b>	0/0.8477	0/0.832
80 %	0/0.778	0/0.778	0/0.778	0/0.778	<b>0/0.780</b>

---

### B4.2.3 ASSEMBLAGE DU GENOME ET ANNOTATION

---

Les séquences d'ADN chloroplastique ainsi obtenues ont été annotées en utilisant le logiciel DOGMA (Dual Organellar GenoMe Annotator (Wyman *et al.*, 2004). Ensuite, celle-ci ont été affinées manuellement en utilisant le logiciel Artemis version 11 (Rutherford *et al.*, 2000). Les limites des codons stop et start pour chaque gène codant pour une protéine, ainsi que les limites des introns et des exons pour les gènes ayant des introns ont été vérifiées manuellement en utilisant comme génome de référence celui de *Castanea mollissima* (NC\_014674). Les ARN de transferts (tRNA) ont été identifiés en utilisant les logiciels DOGMA et tRNAscan-SE (Lowe and Eddy, 1997) avec les paramètres par défauts. La carte circulaire du génome chloroplastique a été dessinée en utilisant le programme OGDRAW version 1.2 (Lohse *et al.*, 2007). Ce travail bio-informatique a été effectué à l'Istituto di Genomica Applicata (IGA), dans le cadre d'une coopération avec le CNR (DR G.G. Vendramin) sur les espèces ayant servies de tests préliminaires.

Une carte du génome chloroplastique (Jansen *et al.*, 2011) basée sur trois espèces (*Castanea mollissima*, *Theobroma cacao* et *Prunus persica*) qui appartiennent au super-ordre des rosidés est représentée dans la figure B16.

---

### B4.2.4 RECONSTRUCTIONS PHYLOGENETIQUES

---

Comme pour les parties précédentes de ce manuscrit, plusieurs types de reconstructions phylogénétiques ont été réalisées (Voir figure B17). La première est basée sur l'ensemble de l'information obtenue suite à l'amplification de l'ADN chloroplastique total sur une SuperMatrice de 57 individus (espèces) \* 164928pb. Pour la suite, et afin de voir s'il est possible de n'utiliser qu'une partie de l'information génétique (deuxième objectif), un sous ensemble des 164928pb a été utilisé.

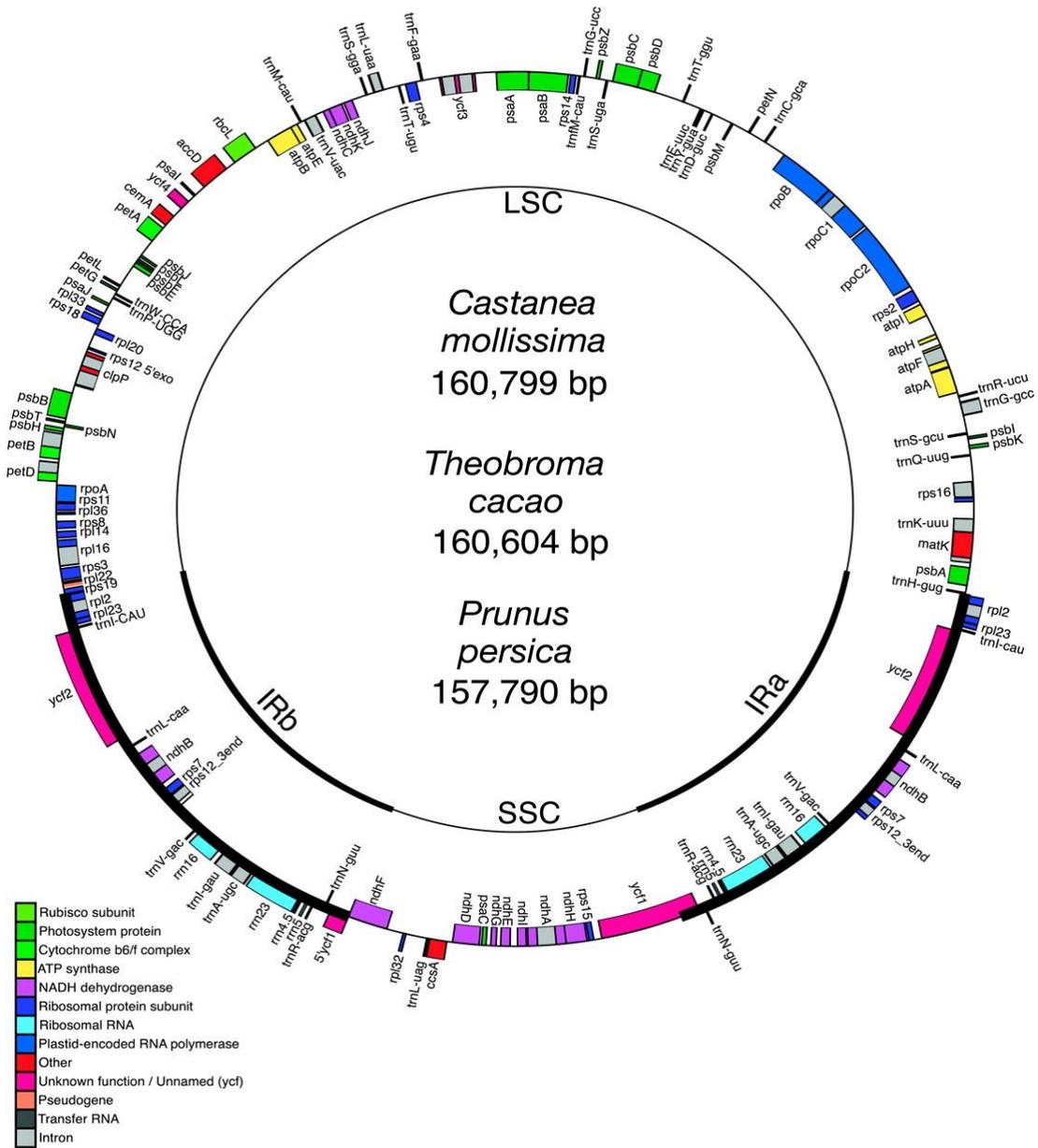


FIGURE B16 : CARTE CIRCULAIRE DU GENOME CHLOROPLASTIQUE DE TROIS ESPECES APPARTENANT AU SUPER ORDRE DES ROSIDES.

#### B4.2.4.1 SUR LE GENOME COMPLET

---

Deux types de reconstructions phylogénétiques ont été menées sur la matrice totale (57 individus \* 164928pb) : une par Maximum de Vraisemblance (MV) avec le logiciel RAxML 7.2.8-ALPHA (Stamatakis, 2006; Stamatakis *et al.*, 2008) et une par Inférence Bayésienne (BI) avec le logiciel MrBayes 3.1.2 (Ronquist and Huelsenbeck, 2003). La recherche par Maximum de Vraisemblance suit les recommandations de l'auteur, à savoir que la recherche se fait en optimisant le modèle d'évolution GTR + Gamma sur chaque gène de façon indépendante, avec 4 types différents de vitesse d'évolution des sites. 1000 re-échantillonnages par bootstrap sont menés en parallèle.

Le nombre de sites polymorphes ainsi que le nombre de sites parcimonieux sont indiqués dans le tableau B12. Y figurent également, le nombre de patterns retournés par le logiciel PhyML v.3.0 (Guindon and Gascuel, 2003) ainsi que le nombre de sites sans polymorphisme.

L'Inférence Bayésienne menée consiste en deux runs parallèles de 10 millions de générations avec échantillonnage tous les 1000 pas. Chaque run est constitué de deux types de chaînes : trois chaînes chaudes et une froide. La convergence entre les deux runs est achevée lorsque la déviation standard moyenne entre les runs est inférieure à 0.05.

Les figures B18 et B19 correspondent à la topologie obtenue lors de l'inférence Bayésienne. Les chiffres indiqués sur les branches correspondent aux valeurs de bootstraps supérieurs à 60% avec les probabilités postérieures associées ; la différence entre les deux provient de la coloration des noms d'espèces : suivant la taxonomie (Figure B18) ou suivant la zone géographique d'origine de l'espèce (Figure B19).

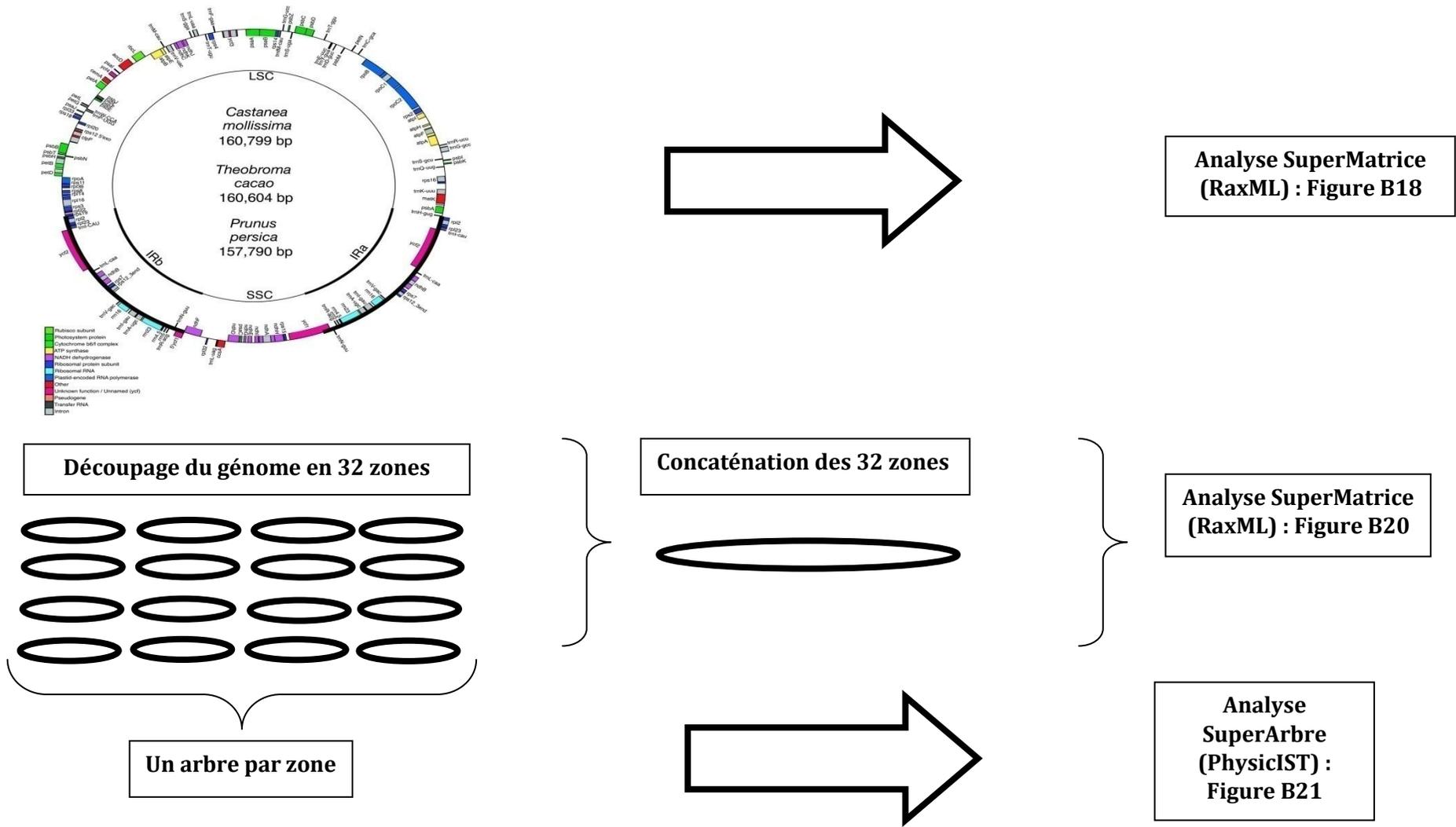


FIGURE B17 : METHODOLOGIE SUIVIE POUR LES DIFFERENTS TYPES DE RECONSTRUCTIONS PHYLOGENETIQUES (sur le génome global et sur les 32 zones).

#### B4.2.4.2 SUR LE SOUS ENSEMBLE

---

L'annotation n'ayant pas été disponible pour toutes les espèces, celle-ci a été transposée du génome chloroplastique du châtaignier (*Castanea mollissima*) sur chaque séquence (espèce) afin de définir différentes zones codantes et non codantes. Ces zones correspondent soit à des gènes codant pour les protéines (dont certains sont traditionnellement utilisés en phylogénie comme MatK, rBcl), soit à des régions intergéniques non codantes, soit enfin à des ARN de transfert ou ribosomiaux (Tableau B12). Un arbre phylogénétique par zone chez toutes les espèces a été produit par Maximum de Vraisemblance par le logiciel PhyML v.3.0 (Guindon and Gascuel, 2003) en appliquant le même modèle d'évolution (HKY + Gamma + Invariants). Le support de chaque branche n'a non pas été réalisé par ré échantillonnage de type bootstrap mais en utilisant le critère aLRT (approximate likelihood-ratio test) qui donne une valeur de soutien aux branches qui va dans le même sens que le bootstrap lorsque la valeur est forte (Anisimova and Gascuel, 2006). Ce critère varie entre 0 et 1 : 0 indique un soutien de branche nul, alors que 1 indique un soutien important. Une zone a été retenue si la proportion de branches dont l'aLRT est supérieure à 0,6 est plus élevée que 20%. En se basant sur ce critère, 32 zones ont été retenues (Tableau B12). Au total ces 32 zones représentent 52745 pb soit 32% du génome chloroplastique.

Pour l'analyse combinée (concaténation des 32 marqueurs), la concaténation des séquences de tous les zones a été faite avec le logiciel Seaview (version 4); puis une reconstruction phylogénétique a été menée par Maximum de Vraisemblance (MV) avec le logiciel RAxML 7.2.8-ALPHA (Stamatakis, 2006). La figure B20 représente cet arbre avec les valeurs de bootstraps supérieurs à 50%.

La méthode SuperArbre combine des arbres phylogénétiques avec des échantillons taxonomiques se recouvrant au moins partiellement entre les arbres phylogénétiques étudiés. Des conflits topologiques peuvent apparaître entre les différents arbres sources pour des raisons méthodologiques ou biologiques comme l'attraction des longues branches ou encore l'hybridation (Scornavacca *et al.*, 2008). Les conflits topologiques sont représentés par les différents arrangements du même taxa entre différents arbres sources. Lorsque des conflits topologiques apparaissent entre les différents arbres sources, il est possible de définir deux types de méthodes de SuperArbre en fonction de comment la méthode de construction de celui-ci les gère : la méthode dite libérale, et la méthode dite de véto.



Les méthodes dites libérales infèrent des SuperArbres contenant l'alternative la plus fréquente. Ceci s'effectue en demandant aux arbres sources de voter et d'opter pour l'alternative maximisant un critère d'optimisation défini (Scornavacca *et al.*, 2008). Dans ce cas, il faut espérer que chaque taxon qui est mal placé l'est seulement pour quelques arbres et que cette erreur soit surpassée par un nombre d'arbre plus important où le taxon est bien placé.

Les méthodes dites de véto, rejettent tous les arbres contenant des informations conflictuelles.

Lorsque les arbres sources entre souvent en conflits topologiques entre eux ou qu'il y a peu de taxa en communs, les SuperArbres issus de ces deux méthodes sont très peu résolutifs (Scornavacca *et al.*, 2008).

Pour dépasser ce problème, la méthode issue du logiciel PhySIC-IST (Scornavacca *et al.*, 2008) disponible sur le lien suivant : <http://www.atgc-montpellier.fr/physic-ist/> a été utilisée. Cette méthode est une variante des méthodes de véto et permet d'éliminer de l'analyse les individus qui provoquent le plus de conflits entre les arbres sources. L'informativité du SuperArbre est calculée par une variante du CIC (Cladistic Information Content). Ce critère prend en compte la présence de multifurcations et l'absence de certains taxa. Cette valeur est définie comme étant la proportion du nombre d'arbres binaires entièrement résolus qui sont compatibles avec le SuperArbre construit. Plus la valeur de CIC est proche de 1, et plus le SuperArbre représente de façon adéquate les informations topologiques des différents arbres sources (Scornavacca *et al.*, 2008). Ce logiciel propose une étape de préprocessing appelée STS (Source Tree Correction) qui corrige les arbres sources avant l'inférence du SuperArbre. Cette étape fait partie de la catégorie des méthodes libérales dans le sens où elle retire les parties des arbres sources qui rentrent en conflit avec les autres. Cette valeur varie de 0 à 1. Une valeur de 1 indique au logiciel de ne pas éliminer les taxa dont la place est incongruente entre les différents arbres, une valeur de 0 lui disant de les éliminer dès qu'ils apparaissent.

Une recherche de l'arbre le plus vraisemblable pour chaque zone a été menée avec le logiciel PhyML v.3.0 en appliquant le même modèle d'évolution pour chacune des zones (HKY + Gamma + Invariants). Le support des branches est calculé en effectuant 100 ré-échantillonnage par bootstrap.

Une fois chacun des arbres raciné, les arbres sont comparés entre eux pour en dériver un arbre congruent (le SuperArbre). Cette analyse est conduite en deux étapes :

- Tout d'abord les branches à l'intérieur de chaque arbre sont consolidées en définissant un niveau seuil de support. Toutes les branches soutenues par une valeur inférieure sont détruites. Seuls les clades fortement soutenus seront considérés par la méthode de SuperArbre.



- Une valeur de détection d'anomalies (STS, Source Tree Correction) est demandée afin de modifier les arbres sources avant la construction du SuperArbre.

Le tableau B13 contient les différents résultats issus de ces analyses. Les différentes valeurs de bootstrap utilisées sont indiquées en colonne. Les valeurs de STS sont indiquées en ligne.

Le nombre d'individus éliminés et la valeur CIC (Cladistic Information Content) mesurant l'informativité du SuperArbre sont indiqués dans les cases. Le SuperArbre le plus informatif (optimisation maximale entre la valeur de bootstrap, le nombre d'individus éliminés le plus faible et le CIC le plus élevé) est donné sur la figure B21 (bootstraps de 70% et STS de 0.4).

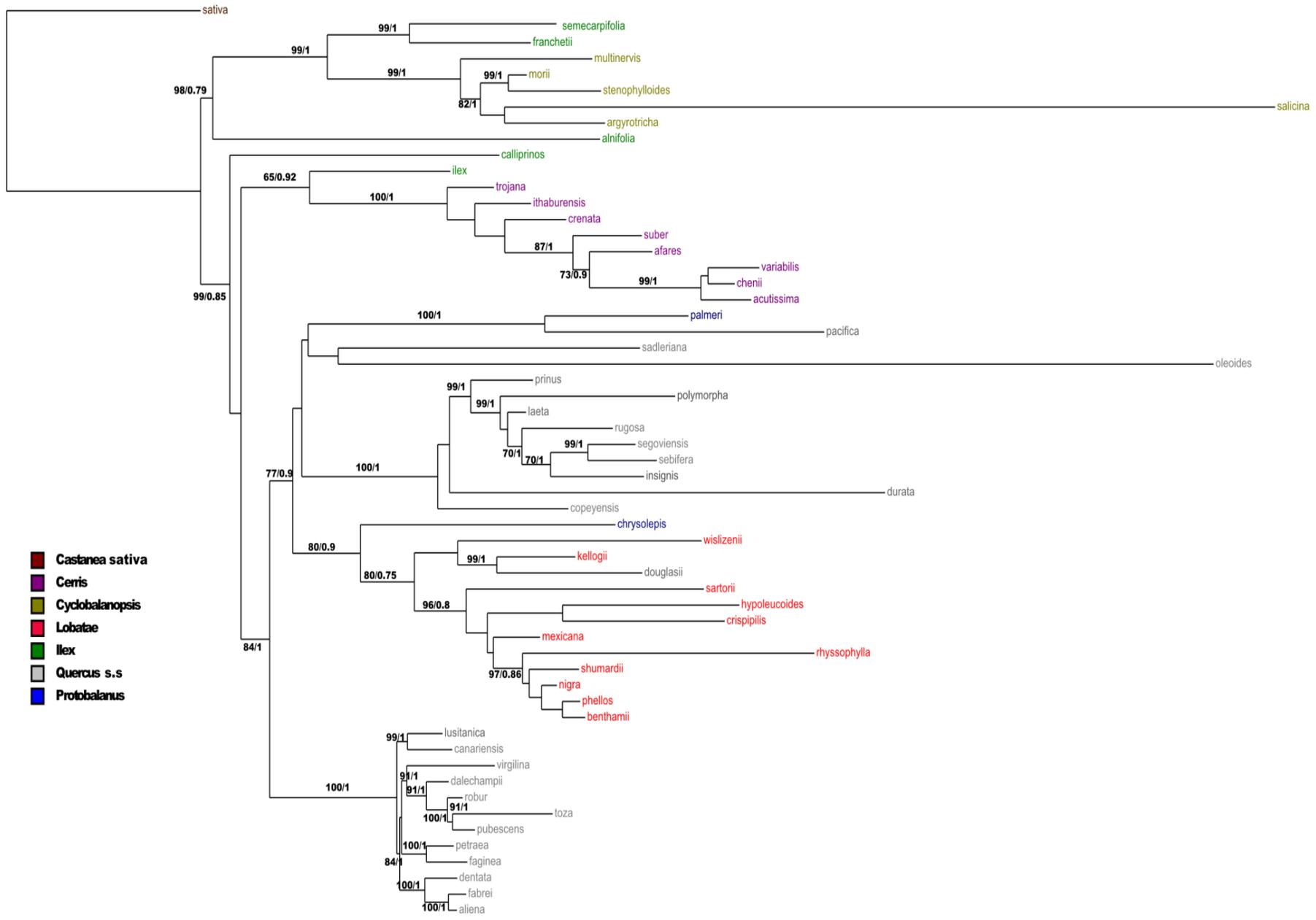


FIGURE B18 : ARBRE PHYLOGENETIQUE BASE SUR LA SEQUENCE TOTALE DU GENOME CHLOROPLASTIQUE. LES VALEURS DE BOOTSTRAPS SUPERIEURES A 60% SONT INDIQUEES AVEC LES PROBABILITES POSTERIEURES ASSOCIEES. LES COULEURS CORRESPONDENT AUX GROUPES TAXONOMIQUES.

## B4.3 RESULTATS

---

### B4.3.1 RESULTAT SUPERMATRICE SUR LE GENOME TOTAL

---

Le nombre de sites polymorphes sur le génome total (Tableau B12) est de 1209 soit un taux de polymorphisme de  $(1209/164928) = 0.007$  avec un nombre de pattern de 20460.

La matrice 57 individus \* 164928 pb a servi de base à la recherche de l'arbre le plus vraisemblable et à l'inférence bayésienne (Figure B18).

La topologie globale est très bien soutenue, avec des fortes valeurs de bootstraps et de probabilités postérieures associées très élevées (Figure B18).

En partant de la racine (*Castanea sativa*), un premier clade composé de toutes les espèces de *Cyclobalanopsis*, et de trois espèces du groupe *Ilex* (*Q. franchetii*, *Q. semicarpifolia* et *Q. alnifolia*) s'individualise. Ce clade est fortement soutenu, avec une valeur de bootstrap de 98% et une probabilité postérieure de 1 (Figure B18) ; nous l'appellerons Groupe A.

Une espèce du groupe *Ilex* (*Q. calliprinos*) se distingue ensuite mais avec un très faible support (Figure B18), puis un groupe (Groupe B) constitué d'une espèce du groupe *Ilex* (*Q. ilex*) et un clade constitué de toutes les espèces du groupe *Cerris*. Ce groupe B est frère d'un clade comprenant les trois autres groupes (*Quercus s.s.*, *Protobalanus* et *Lobatae*). Ce clade est lui même scindé en deux : un sous-groupe de *Quercus s.s.* (Groupe C) très fortement soutenu (bootstraps de 100 % et probabilité postérieure de 1) et exclusivement composée d'espèces eurasiatiques de *Quercus s.s.*, et du Groupe D composé d'espèces exclusivement nord américaines représentant (*Quercus s.s.*, *Protobalanus* et *Lobatae*). Les deux groupes C et D sont eux aussi très fortement soutenus. Cet arbre se caractérise globalement par une bonne résolution, l'éclatement géographique de certains groupes taxonomiques (notamment le groupe *Ilex* et le groupe *Quercus s.s.*), et l'assignation surprenante des espèces du groupe *Protobalanus* soit à *Quercus s.s.* (*Q. palmeri*) soit à *Lobatae* (*Q. chrysolepis*).

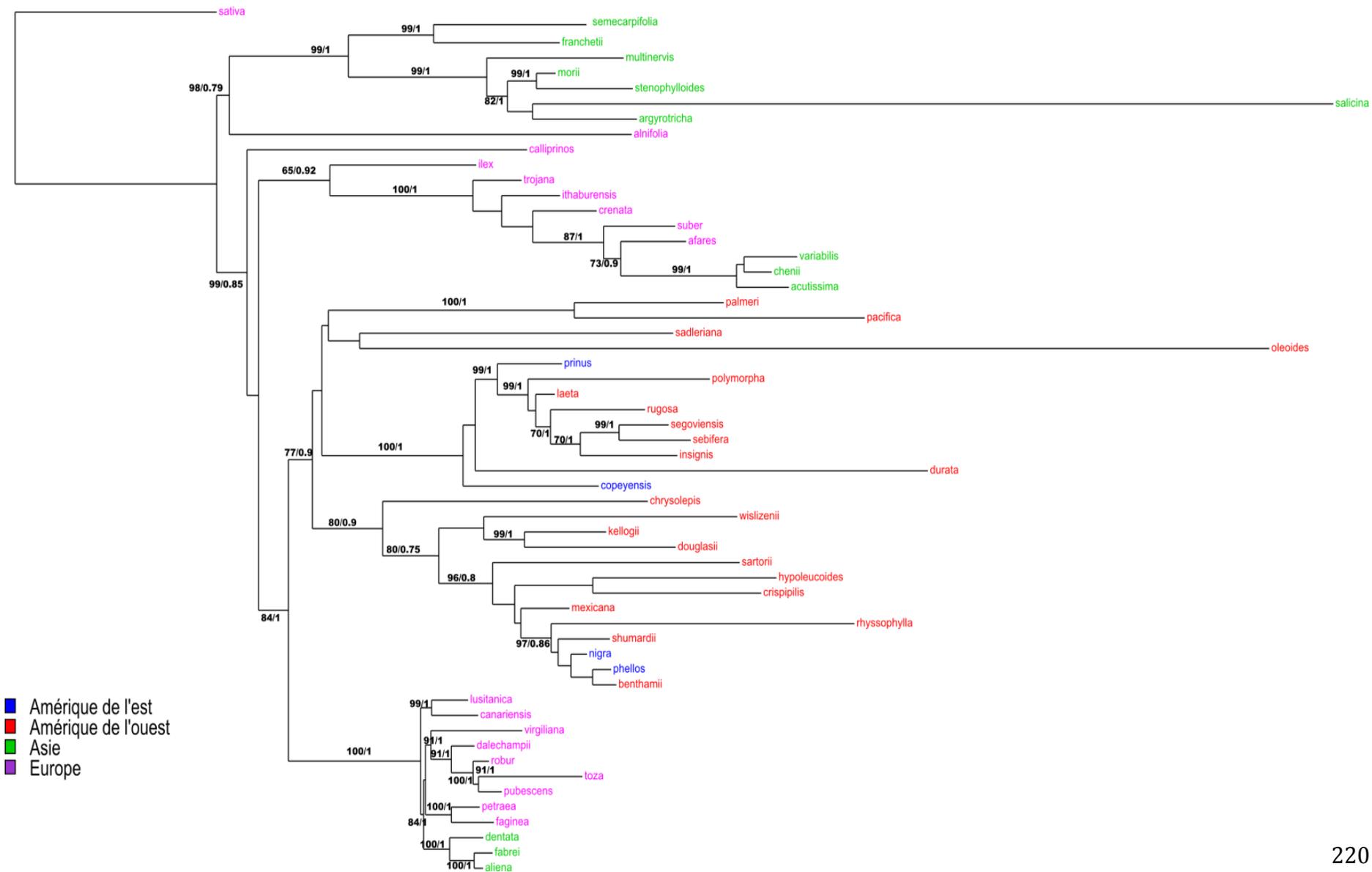


FIGURE B19 : ARBRE PHYLOGENETIQUE BASE SUR LA SEQUENCE TOTALE DU GENOME CHLOROPLASTIQUE. LES VALEURS DE BOOTSTRAPS SUPERIEURES A 60% SONT INDIQUEES AVEC LES PROBABILITES POSTERIEURES ASSOCIEES. LES COULEURS CORRESPONDENT AUX QUATRES ZONES GEOGRAPHIQUES.

---

### B4.3.2 RESULTATS SUPERMATRICE SUR LE SOUS ENSEMBLE

---

Les 32 zones utilisées comme marqueurs (Tableau B12) sont caractérisées par une longueur variable de 351 et 5701 pb et un polymorphisme moyen (nombre de sites variables / nombre de sites total) de 0.011.

L'arbre phylogénétique issu des 32 gènes concaténés en une seule SuperMatrice retrouve les 4 groupes préalablement définis (Figure B20), mais la topologie de l'arbre n'est pas la même. Le premier groupe ayant divergé après la racine (*Castanea sativa*), étant le groupe A avec une valeur de support de branche élevée (70 %, Figure B20). Le groupe ayant divergé ensuite n'est non pas le groupe B comme précédemment (Figure B19) mais les 3 groupes ensemble formant une polytomie, et ce, de façon soutenue (bootstraps de 98 %, Figure B19). Le groupe A est frère d'un super groupe comprenant les groupes B, C et D (topologie A-(B+C+D)), alors que la topologie précédente était (A-B-(C+D)). Le SuperArbre apparaît donc globalement moins résolu. Au-delà de sa moins bonne résolution, l'arbre de la Figure B20 se caractérise par une autre particularité : la réunion du groupe des *Quercus s.s.* européens et du groupe B (*Cerris +Ilex*) en un clade frère du clade nord-américain constitué des espèces du *Quercus s.s.* (espèces Nord-Américaines), et des *Lobatae* et *Protobalanus*). Néanmoins la branche du clade européen est faiblement soutenue (bootstrap inférieur à 50%, Figure B20).

---

### B4.3.3 RESULTAT SUPERARBRE SUR LE SOUS ENSEMBLE

---

Les différentes combinaisons entre les valeurs de bootstraps (40, 50, 60, 70 et 80 %) et la valeur de la correction STC (0.1, 0.2, 0.4, 0.6, 0.8) sont indiquées dans le tableau B13. Ces seuils ont été choisis dans le but de balayer au mieux la diversité de ces critères. Le logiciel décide d'enlever entre zéro et deux individus de l'analyse, avec une valeur de significativité CIC allant de 0.778 à 0.924 (Tableau B13). Au regard de ces résultats, le SuperArbre produit reflète bien les informations topologiques des arbres sources (CIC proche de 1), tout en enlevant très peu d'individus suite à la correction par le paramètre STC (paramètre corrigeant les arbres sources pour éviter les conflits lors de la recherche du SuperArbre le plus représentatif).

Le SuperArbre reconnaît un groupe constitué par les groupes A et B (définis lors de la visualisation de l'arbre issu de la SuperMatrice globale), puis un deuxième groupe constitué des deux autres groupes sans liens entre eux (Figure B16).

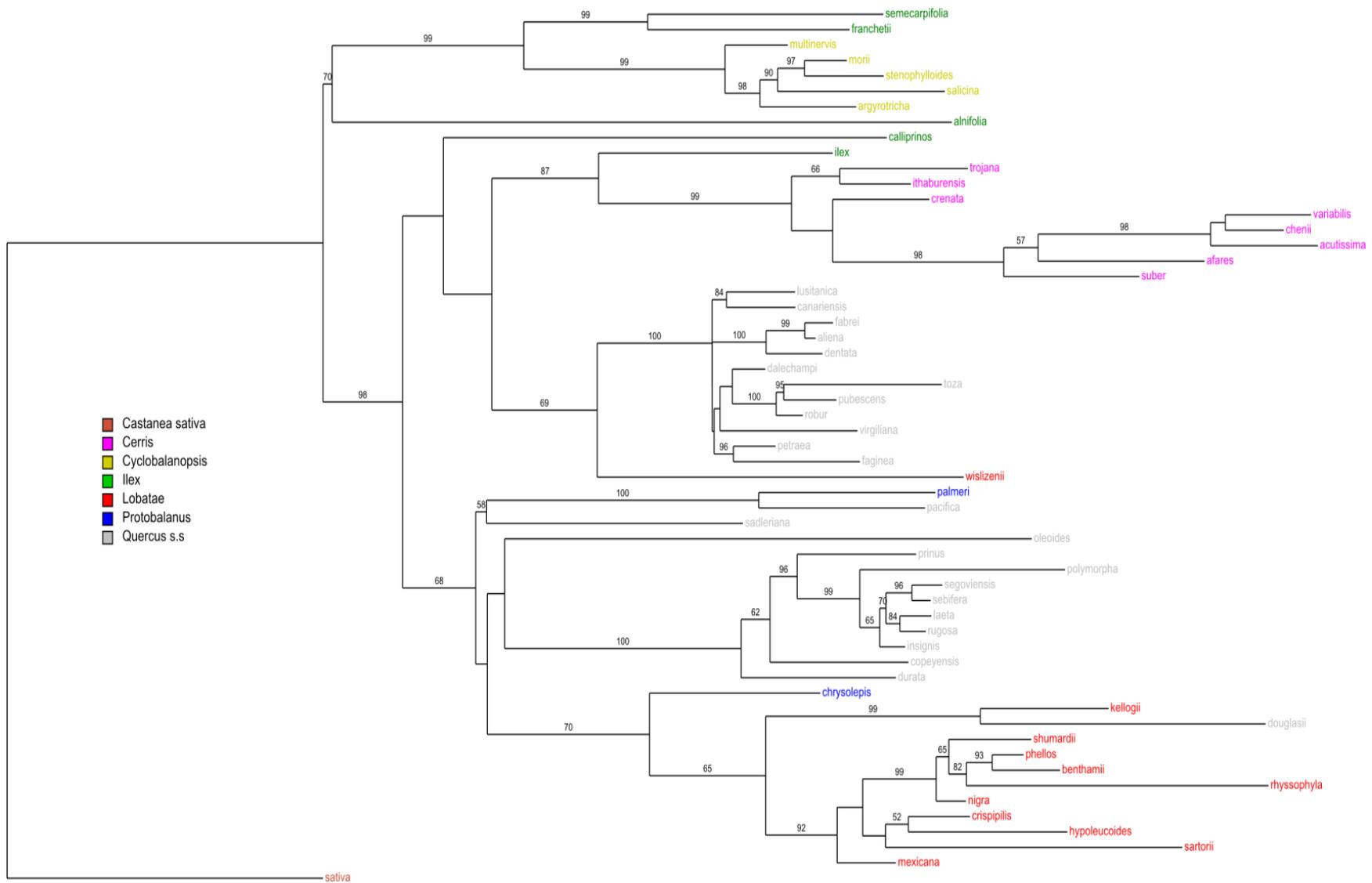


FIGURE B20 : ARBRE PHYLOGENETIQUE ISSU D'UNE SUPERMATRICE BASEE SUR 32 ZONES DU GENOME CHLOROPLASTIQUE. LES VALEURS DE BOOTSTRAPS SUPERIEURES A 50% SONT INDIQUEES.

## B4.4 DISCUSSION/CONCLUSION

---

L'objectif premier de cette étude était de voir si le séquençage total (164928 pb) du génome chloroplastique sur une soixantaine d'espèces serait capable de donner un arbre phylogénétique plus résolu avec des branches plus soutenues que dans les précédentes études. Ce premier objectif a été atteint comme en témoigne la figure B18 ; en effet, tous les clades sont caractérisés par des branches soutenues, et un niveau de résolution élevé de cet arbre. La comparaison de cet arbre avec les arbres issus des données nucléaires sera discutée ultérieurement. Nous nous limitons ici à la description des résultats issus du séquençage du génome chloroplastique et nous mettrons ceci en lien avec les quelques études précédentes de ce même génome.

Notre analyse a également mis en évidence la « perte » de signal phylogénétique quand les reconstructions phylogénétiques sont limitées à une portion plus réduite du génome. L'analyse portant sur les 32 zones « extraites » du génome (soit 32%) génère un arbre moins résolu et dont les branches sont caractérisées par des supports plus faibles. L'approche SuperMatrice (concaténation des 32 zones, Figure B20) et l'approche SuperArbre (comparaison de la topologie de chaque arbre phylogénétique, Figure B21) aboutissent dans les deux cas à des niveaux de résolutions inférieurs à celui obtenu avec la séquence complète de l'ADN chloroplastique. Il y a donc bien une érosion du signal phylogénétique qui est diffus, associée à une réduction de la prospection du génome chloroplastique.

Enfin, le dernier objectif de cette partie portait sur l'empreinte phylogéographique laissée dans l'ADN chloroplastique. Nos résultats montrent une surimposition d'une structure géographique sur la structure phylogénétique issue de l'analyse des gènes nucléaires. La conjonction de ces deux structures est particulièrement illustrée par la position du groupe des chênes blancs (*Quercus s.s.*) sur l'arbre de la figure B18. Rappelons que ce groupe est le seul dont les espèces actuelles sont distribuées sur les deux continents, et était apparu comme un clade unique dans l'arbre phylogénétique construit sur les gènes nucléaires (voir la figure B11 de la partie B3 du manuscrit).

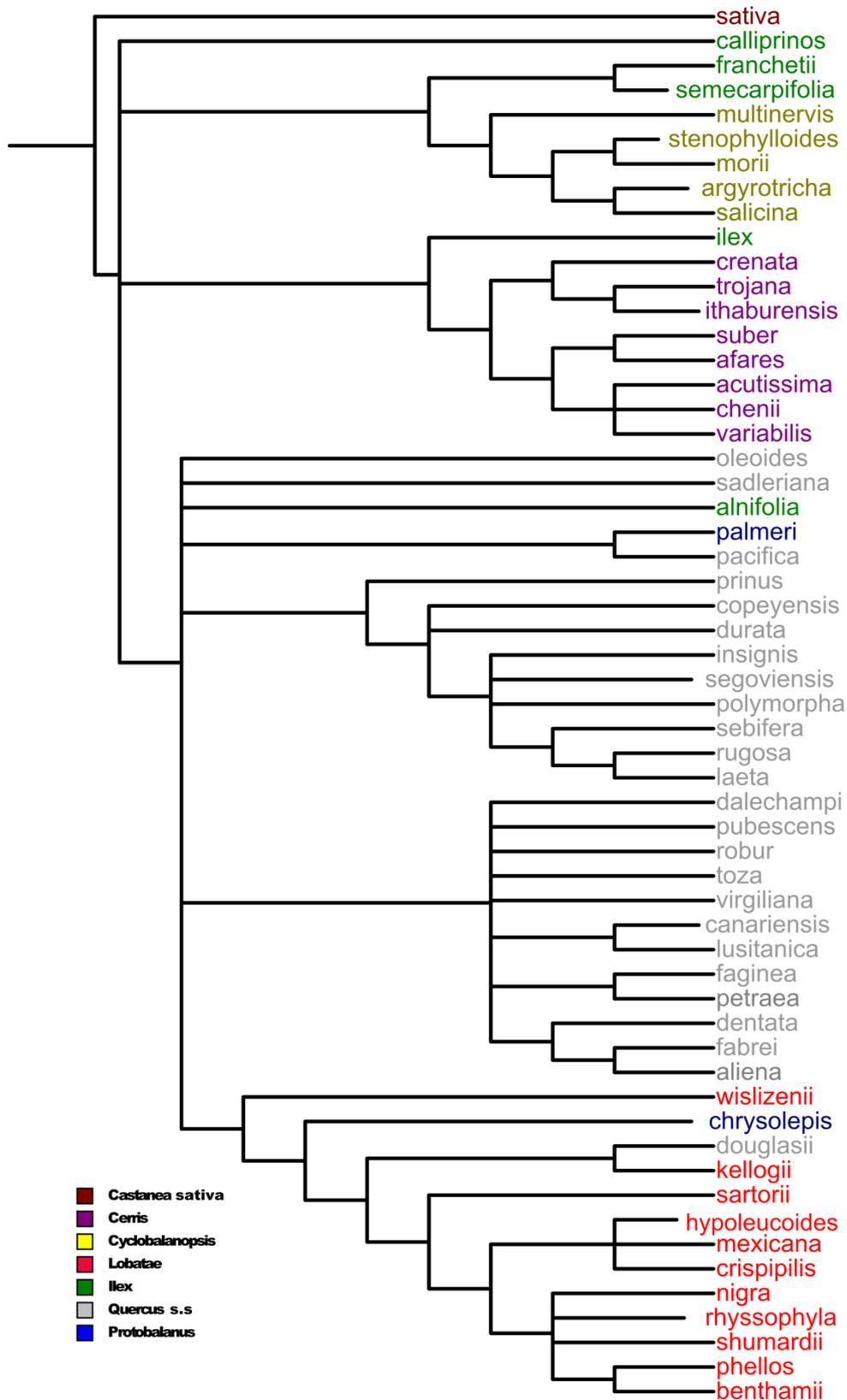


FIGURE B21 : SUPERARBRE BASE SUR 32 ZONES DU GENOME CHLOROPLASTIQUE.

Ce clade est éclaté dans la phylogénie chloroplastique en deux sous clades correspondant aux espèces eurasiatiques et américaines (Figure B18 et B19). Par ailleurs, le groupe Eurasiatique lui-même se scinde en un groupe asiatique et européen, et le groupe européen en espèces occidentales (*Q. lusitanica* et *Q. canariensis*) et espèces d'Europe centrale. Le clade des *Cerris* (Figure B18 et B19), qui n'est constitué que d'espèces eurasiatiques, est éclaté en espèces asiatiques et européennes. Le signal géographique de l'ADN chloroplastique qui avait été initialement mis en évidence au niveau infraspécifique et microévolutif (Petit *et al.*, 2002), se trouve ici confirmé au niveau générique et macroévolutif. Comme indiqué précédemment, ces résultats confirment ce qui était déjà perceptible dans l'analyse du fragment DT de l'ADN chloroplastique (Xu, 2004). Manos (Manos *et al.*, 1999) quant à lui, en utilisant l'ensemble de l'information issue des sites de restrictions de l'ADN chloroplastique n'est pas arrivé à ce résultat (Figure A10), sans doute du fait d'un sous échantillonnage à l'intérieur des groupes taxonomiques.

Les résultats obtenus au niveau de l'ADN chloroplastique nous permettent d'émettre deux grandes conclusions, au niveau technique et biogéographique.

#### Au niveau technique :

La motivation essentielle qui avait conduit à l'exploration des ressources génomiques dans le cadre de ce travail de thèse était la recherche d'un signal phylogénétique suffisant pour améliorer la résolution de la phylogénie du genre *Quercus*. Tous les travaux antérieurs avaient buté sur la polytomie souvent présente à la base des arbres. Nos résultats obtenus avec l'ADN chloroplastique nous permettent de répondre partiellement à cette question.

Il est évident que ce n'est pas uniquement le nombre de sites polymorphes qui soit le meilleur indicateur de la résolution de la phylogénie, puisque la qualité du signal phylogénétique est primordiale. Mais cette valeur comparée entre les différents gènes ou combinaisons de gènes utilisés peut être un indice nous permettant de conclure sur le fait qu'il soit nécessaire ou non de séquencer un nombre conséquent de nucléotides dans le but d'obtenir une phylogénie résolue.

Le taux de polymorphisme moyen (nombre de sites polymorphes/taille) de l'ADN chloroplastique est très inférieur (0.007) aux taux issus des six gènes nucléaires de la partie B3 (Tableau B7 de la partie B3 du manuscrit) qui à une valeur de 0.19 ; mais également, à celui issu du gène CRC qui à une valeur de 0.32 (Tableau B7 de la partie B3 du manuscrit).

Le nombre de sites polymorphes permettant d'obtenir une phylogénie résolue sur 57 espèces à partir du séquençage chloroplastique est de 1209 (Tableau B12). Cinq cent cinquante sites polymorphes (0.19\*2896 pb) ont été retrouvés dans les six gènes nucléaires issus de cette étude,



sur 112 espèces. Neuf cent quarante six sites polymorphes (0.32\*2957) l'ont été sur 52 espèces du gène CRC (Oh and Manos, 2008). Il faudrait donc bien augmenter le nombre de sites polymorphes sur les gènes nucléaires pour pouvoir espérer atteindre le niveau de résolution issu du séquençage complet de l'ADN chloroplastique.

L'hypothèse selon laquelle le signal phylogénétique serait diffus et conduirait à une apparente mais fausse irrésolution sur le génome nucléaire comme nous l'avons vu dans la partie introductive de ce manuscrit et la deuxième partie, paraît donc prendre toute sa valeur. Il semble donc important de continuer dans la découverte de nouveaux marqueurs nucléaires pour espérer gagner en résolution et en valeur de soutien des branches profondes des phylogénies.

#### Au niveau biogéographique :

La très forte empreinte géographique perceptible dans l'ADN chloroplastique nous permet de préciser les scénarios de biogéographie historique du genre. Rappelons que deux grands scénarios d'évolution du genre *Quercus* avaient été émis dans la littérature à ce jour :

- Le centre de diversification serait situé en Asie du Sud Est, à partir duquel le genre a suivi deux voies migratoires, vers l'Ouest et l'Europe puis l'Amérique du Nord (via une migration par le pont Nord Atlantique) d'une part, et vers l'Est et l'Amérique du Nord (via une migration par le détroit de Béring) d'autre part. Le genre *Trigonobalanus* présent en Indochine aurait donné le sous-genre *Cyclobalanopsis* et le sous-genre *Quercus s.l.*. L'ancêtre du sous-genre *Quercus s.l.* Asiatique aurait divergé en allant vers l'ouest en *Cerris* Asiatique, qui aurait donné naissance au groupe *Ilex* en Europe puis aux *Protobalanus* après passage vers l'Amérique du Nord. Ce même ancêtre du sous-genre *Quercus s.l.* aurait divergé après passage de l'Asie à l'Amérique du Nord via le détroit de Béring en *Quercus s.s.* C'est le scénario de Zhou (Zhou, 1992) ;
- Le genre s'est différencié à partir d'une forêt ancestrale couvrant tout l'hémisphère nord et qui aurait été le départ des deux centres de diversification : l'Asie du Sud Est et l'Amérique du Nord. Le centre Asiatique de diversification aurait donné naissance aux groupes *Cyclobalanopsis*, *Cerris* et *Ilex* et le centre Américain aux groupes *Quercus s.s.*, *Protobalanus* et *Lobatae*. Les chênes blancs (*Quercus s.s.*) auraient migré depuis l'Amérique vers l'Asie et l'Europe C'est l'hypothèse plus récente de Manos et Stanford (Manos and Stanford, 2001), reprise depuis par Denk et Grimm (Denk and Grimm, 2010).



Nos résultats suggèrent que le premier scénario semble plus vraisemblable que le second. Le premier groupe de l'arbre phylogénétique, manifestant les caractères les plus plésiomorphes, comprend en effet des espèces appartenant aux groupes *Cyclobalanopsis* et *Ilex* (Figure B18). Les espèces du groupe *Ilex* sont celles de la section et sous section *Heterobalanus* de Menitski (Menitsky, 1984; Menitsky, 2005), aujourd'hui situées à haute altitude sur le versant est de l'Himalaya (section *Brachylepides* de Camus). La seule espèce européenne affiliée à ce groupe dans la classification de Menitski est *Q. alnifolia*, espèce relique aujourd'hui endémique à Chypre. Or les espèces *Q. pseudoalnus* et surtout *Q. sosnowskyi*, espèces fossiles manifestant de fortes affinités avec *Q. alnifolia*, avaient une très large distribution eurasiatique (Palamarev, 1989; Palamarev and Ivanov, 2003; Menitsky, 2005). Les espèces du groupe *Ilex* sont également celles pour lesquelles les restes fossiles européens les plus anciens remontent à l'Oligocène, alors que les fossiles des espèces des groupes *Cerris* et *Quercus s.s.* remontent au Miocène (Palamarev, 1989). La position des espèces du groupe *Ilex* et ces éléments historiques semblent donc bien corroborer l'hypothèse que le genre *Quercus* aurait dérivé d'un ancêtre portant des caractères *Ilex* (Denk and Grimm, 2009), issu d'Asie et ayant migré vers l'Ouest. Une première diversification aurait conduit au groupe *Cerris* présent en Asie et en Europe. Sur l'arbre phylogénétique chloroplastique, *Cerris* est un clade frère de l'ensemble des trois groupes constituant les espèces du nouveau monde, dont l'ancêtre était sans doute eurasiatique. Deux hypothèses deviennent alors possibles :

- Soit cet ancêtre s'est diversifié en Eurasie en espèces du groupe *Quercus s.s.* dont certaines ont migré en Amérique via le détroit de Béring (ou par l'Atlantique Nord) et y ont donné naissance aux deux autres groupes *Lobatae* et *Protobalanus*.
- Soit cet ancêtre a migré en Amérique du Nord et s'est diversifié sur place en *Quercus*, *Lobatae* et *Protobalanus*. Une migration en retour des *Quercus s.s.*, via le détroit de Béring a permis à ce groupe de coloniser l'Europe.

Dans les deux hypothèses des migrations via le détroit de Béring sont nécessaires et semblent attestés par des restes fossiles de *Quercus* observés en Alaska (Leopold *et al.*, 2007) et en Russie (Fot'yanova, 1985). Les reconstructions phylogénétiques ne nous permettent pas de trancher entre ces deux hypothèses (Figure B18). Des éléments historiques complémentaires sont nécessaires pour conforter l'une ou l'autre hypothèse. Par contre l'origine unique et asiatique du genre semble confortée pour la première fois par des analyses phylogénétiques. De la même manière que pour les analyses nucléaires, nous avons essayé de calibrer la phylogénie basée sur le génome chloroplastique. L'analyse consistant à dater les nœuds avec le logiciel BEAST a été lancée mi mars 2013, et ne donne actuellement pas de résultats : la convergence est très longue à atteindre et une analyse correcte devrait mettre plusieurs mois (le dernier résultat qui a mis deux semaines indique qu'il faudrait multiplier ce temps par au moins 50 !!!)...



## B4.5 DEROULEMENT DE CETTE ETUDE

---

Dans le cadre du projet Européen Evoltree coordonné par Antoine Kremer, le séquençage du génome chloroplastique total a été réalisé au tout début de ma thèse sur quelques espèces. Il a donc été convenu que toutes les espèces utilisées pour les phylogénies nucléaires le soient également au niveau du génome chloroplastique. Cette sous-traitance s'est réalisée en collaboration avec Beppe Vendramin qui travaille à l' « Institute of Plant Genetics » à Florence (Italie) et partenaire du réseau Evoltree C'est dans ce cadre que je suis parti une semaine en Italie en Janvier 2012 avec quelques échantillons afin de participer aux traitements effectués sur ces échantillons en vue du séquençage. Les dernières séquences ont été reçues fin Janvier 2013, augmentant l'échantillonnage à 57 individus (espèces).

Les cartes de distributions des espèces ont été produites par Paul Fromage, actuellement en contrat à l'OAFS (Observatoire Aquitain de la Faune Sauvage) de l'Université Bordeaux 1.

Des tests bioinformatiques ont été menés quelques mois auparavant au sein de notre unité par Jean-Marc Frigerio et Alain Franc dans le but de voir si la recherche de zones portant du signal phylogénétique pouvait être appliquée sur un jeu de données important vue la taille des séquences chloroplastiques. Ce travail ayant rendu possible une bonne partie de la présente étude.

L'analyse Bayésienne par MrBayes a été effectuée et optimisée par Emmanuelle Jouselin du CBGP.

L'analyse du génome chloroplastique a failli ne pas pouvoir être incluse dans ce manuscrit, car ma thèse ayant débutée en janvier 2010, aurait dû finir fin décembre 2012.

Idéalement, cette partie aurait dû se faire sous la forme d'un article en anglais mais cela se fera à la suite de la thèse; la rédaction de cette partie ayant débutée début Avril 2013.



## B4.6 REFERENCES BIBLIOGRAPHIQUES

---

- Anisimova, M. and O. Gascuel (2006). "Approximate Likelihood-Ratio Test for Branches: A Fast, Accurate, and Powerful Alternative." Systematic Biology **55**(4): 539-552.
- Bendich, A. J. (1987). "Why Do Chloroplasts and Mitochondria Contain So Many Copies of Their Genome." Bioessays **6**(6): 279-282.
- Curtis, S. E. and M. T. Clegg (1984). "Molecular Evolution of Chloroplast DNA-Sequences." Molecular Biology and Evolution **1**(4): 291-301.
- Denk, T. and G. W. Grimm (2009). "Significance of Pollen Characteristics for Infrageneric Classification and Phylogeny in *Quercus* (Fagaceae)." International Journal of Plant Sciences **170**(7): 926-940.
- Denk, T. and G. W. Grimm (2010). "The oaks of western Eurasia: Traditional classifications and evidence from two nuclear markers." Taxon **59**(2): 351-366.
- Dumoulin, S., *et al.* (1995). "Inheritance of chloroplast and mitochondrial genomes in pedunculate oak investigated with an efficient PCR method." Theoretical and Applied Genetics **91**: 1253-1256.
- Edgar, R. C. (2004). "MUSCLE: multiple sequence alignment with high accuracy and high throughput." Nucleic Acids Research **32**(5): 1792-1797.
- Fot'yanova, L. I. (1985). "The seldovian regional stage in Alaska and its analogs in the soviet Far east. ." International Geology Review(27): 669-677.
- Gaut, B. S. (1998). "Molecular clocks and nucleotide substitution rates in higher plants." Evolutionary Biology, Vol 30 **30**: 93-120.
- Govaerts, R. and D. G. Frodin (1998). World Checklist and Bibliography of Fagales (Betulaceae, Corylaceae, Fagaceae and Ticodendraceae).
- Grivet, D., *et al.* (2001). "Genome walking with consensus primers: application to the large single copy region of chloroplast DNA." Molecular Ecology Notes **1**(4): 345-349.
- Guindon, S. and O. Gascuel (2003). "A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood." Systematic Biology **52**(5): 696-704.
- Hughes, C. E., *et al.* (2006). "From famine to feast? Selecting nuclear DNA sequence loci for plant species-level phylogeny reconstruction." Philosophical Transactions of the Royal Society B-Biological Sciences **361**(1465): 211-225.



- Jansen, R. K., *et al.* (2011). "Complete Plastid Genome Sequences of Three Rosids (Castanea, Prunus, Theobroma): Evidence for At Least Two Independent Transfers of *rpl22* to the Nucleus." *Molecular Biology and Evolution* **28**(1): 835-847.
- Kremer, A., *et al.* (2012). "Genomics of Fagaceae." *Tree Genetics & Genomes* **8**(3): 583-610.
- Leopold, E., *et al.* (2007). "Overview of Alaskan tertiary flora-Building on the work of Jack Wolfe." *Cour. Forch.-Inst. Senckenberg* **258**: 129-138.
- Lohse, M., *et al.* (2007). "OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes." *Current Genetics* **52**(5-6): 267-274.
- Lowe, T. M. and S. R. Eddy (1997). "tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence." *Nucleic Acids Research* **25**(5): 955-964.
- Manos, P S., *et al.* (1999). "Phylogeny, Biogeography, and Processes of Molecular Differentiation in Quercus Subgenus Quercus (Fagaceae)." *Molecular Phylogenetics and Evolution* **12**(3): 333-349.
- Manos, P. S. and A. M. Stanford (2001). "The Historical Biogeography of Fagaceae: Tracking the Tertiary History of Temperate and Subtropical Forests of the Northern Hemisphere." *International Journal of Plant Sciences* **162**(S6): S77-S93.
- Menitsky, J. L. (1984). *Duby Asii*. Leningrad: Nauka. .
- Menitsky, Y. L. (2005). *Oaks of Asia*, Science Publishers.
- Oh, S. H. and P. S. Manos (2008). "Molecular phylogenetics and cupule evolution in Fagaceae as inferred from nuclear CRABS CLAW sequences." *Taxon* **57**(2): 434-451.
- Palamarev, E. (1989). "Paleobotanical Evidences of the Tertiary History and Origin of the Mediterranean Sclerophyll Dendroflora." *Plant Systematics and Evolution* **162**(1-4): 93-107.
- Palamarev, E. and D. Ivanov (2003). "A contribution to the Neogene history of Fagaceae in the Central Balkan area." *Acta Palaeobot.* **43**: 51-59.
- Perry, A. S. and K. H. Wolfe (2002). "Nucleotide substitution rates in legume chloroplast DNA depend on the presence of the inverted repeat." *Journal of Molecular Evolution* **55**(5): 501-508.



- Petit, R. J., *et al.* (2002). "Identification of refugia and post-glacial colonisation routes of European white oaks based on chloroplast DNA and fossil pollen evidence." Forest Ecology and Management **156**(1-3): 49-74.
- Ronquist, F. and J. P. Huelsenbeck (2003). "MrBayes 3: Bayesian phylogenetic inference under mixed models." Bioinformatics **19**(12): 1572-1574.
- Rutherford, K., *et al.* (2000). "Artemis: sequence visualization and annotation." Bioinformatics **16**(10): 944-945.
- Scornavacca, C., *et al.* (2008). "PhySIC\_IST: cleaning source trees to infer more informative supertrees." Bmc Bioinformatics **9**.
- Stamatakis, A. (2006). "RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models." Bioinformatics **22**(21): 2688-2690.
- Stamatakis, A., *et al.* (2008). "A Rapid Bootstrap Algorithm for the RAxML Web Servers." Systematic Biology **57**(5): 758-771.
- Wolfe, K. H., *et al.* (1987). "Rates of Nucleotide Substitution Vary Greatly among Plant Mitochondrial, Chloroplast, and Nuclear Dnas." Proceedings of the National Academy of Sciences of the United States of America **84**(24): 9054-9058.
- Wyman, S. K., *et al.* (2004). "Automatic annotation of organellar genomes with DOGMA." Bioinformatics **20**(17): 3252-3255.
- Xu, L. (2004). Diversité de l'ADN chloroplastique et relations phylogénétiques au sein des Fagacées et du genre *Quercus*. Thèse de l'Université Henri Poincaré ,Nancy, France.
- Zhou, Z.-K. (1992). "Origin, phylogeny, and dispersal of *Quercus* from China." Acta Bot Yunnanica **14**: 227-236.



B5 COMPARAISON ENTRE LE GENOME NUCLEAIRE ET LE  
GENOME CHLOROPLASTIQUE : APPORTS A LA BIOGEOGRAPHIE  
HISTORIQUE DU GENRE QUERCUS

---



La phylogénie moléculaire nucléaire établie à partir de six nouveaux marqueurs moléculaires (Partie B3 du présent manuscrit) n'a pas été capable de résoudre totalement les relations entre les six groupes (Figures B10 et B11). Cependant, ces six marqueurs ont été capables d'établir une relation très soutenue entre les *Quercus s.s* et les *Protobalanus* (Figures B10 et B11). La comparaison entre la présente étude et les études précédentes basées sur des inférences phylogénétiques et/ou sur les études morphologiques nous a permis de dessiner une phylogénie hypothétique (Figure 4). Cette phylogénie reconnaît le groupe des *Cyclobalanopsis* comme étant un sous-genre à part entière, les cinq autres groupes appartenant au sous-genre *Quercus s.l* (Figure 4). Cette phylogénie n'est pas totalement résolue : les groupes *Cerris* et *Ilex* étant associés en groupes frères avec une faible valeur de soutien (Figures B10 et B11). Partant du principe que ceci est dû à un manque d'espèces du groupe *Ilex* asiatique, il ne nous est pas possible à ce stade de définir l'antériorité d'un groupe par rapport à l'autre. Le groupe des *Protobalanus* et des *Quercus s.s* est, quant à lui, très bien soutenu (Figures B10 et B11) ; il ne sera donc pas remis en question à ce stade.

L'utilisation du génome chloroplastique entier (Partie B4 du présent manuscrit) a permis de gagner en résolution globale au niveau de la phylogénie qui en découle (Figure B18). Cette phylogénie regroupe les *Cyclobalanopsis* et les *Ilex* asiatique, tout en sachant que cela pourrait provenir d'un sous-échantillonnage taxonomique de ces groupes donnant une irrésolution. Nous pouvons cependant affirmer que le groupe *Ilex* est antérieur au groupe *Cerris* (Figure B18). Sur cette figure, les *Ilex* asiatiques apparaissent antérieurs aux *Ilex* européens (Figure B19) : on pourrait y discerner une diversification vers l'Ouest à mesure que les *Ilex* ont migré vers l'Europe. De même, les *Cerris* européens ont probablement migré en Asie (Figure B19).

Une espèce de *Protobalanus* est associée aux *Lobatae* et la seconde au groupe des *Quercus s.s* (Figure B18). Ces placements inattendus peuvent être liés à des captures du génome cytoplasmique induits par l'hérédité maternelle de ces génomes, et qui ont pu avoir cours dans le cas d'hybridation interspécifique entre des espèces de groupes différents (*Protobalanus* et *Quercus s.s* d'une part, et *Protobalanus* et *Lobatae* d'autre part). La phylogénie basée sur le génome nucléaire a associé très fortement les *Quercus s.s* aux *Protobalanus*. Malgré le placement inattendu de deux espèces des *Protobalanus*, nous garderons l'association entre les *Protobalanus* et les *Quercus s.s* comme fidèle à l'évolution du groupe.

Les *Quercus s.s* américains et eurasiatiques sont séparés d'après la phylogénie chloroplastique alors qu'ils sont regroupés lors de l'inférence phylogénétique basée sur le génome nucléaire (Figures B10 et B11) ; ceci provient de l'empreinte géographique de l'ADN chloroplastique qui indique donc une histoire différente liée à la géographie. Il y a certainement eu une migration



des *Quercus s.s* d'un continent vers l'autre. Il n'est pas possible que les *Quercus s.s* aient une origine européenne puis que leur ancêtre ait migré en Amérique pour donner les *Protobalanus* et les *Lobatae* comme suggéré dans la partie B4. En effet, ceci conduirait à un regroupement des *Protobalanus* avec les *Lobatae* et non pas avec les *Quercus s.s* comme fortement soutenu par la phylogénie nucléaire.

L'intégration de toutes ces considérations nous conduit à redéfinir une phylogénie hypothétique (Figure B22) qui mérite d'être infirmée ou confirmée dans le futur en :

- 1) augmentant l'échantillonnage taxonomique du groupe des *Ilex* Asiatiques,
- 2) séquençant le génome chloroplastique entier et les six marqueurs nucléaires de tous les individus manquant à l'une ou l'autre des deux phylogénies obtenues,
- 3) recherchant de nouveaux marqueurs nucléaires ou séquençant ceux de la littérature (CRABS CLAW par exemple).

Deux scénarios de biogéographie historique du genre mettant à la fois en jeu des phénomènes de dispersion et de vicariance ou uniquement de dispersion peuvent ainsi être dégagés si on considère uniquement la séparation des continents européens et américains comme moteur de la spéciation.

En considérant que la séparation Europe/Amérique (ou la liaison via le détroit de Bering et la Fennoscandinavie) soit le moteur principal de l'évolution de ce genre, alors l'origine de *Quercus* serait asiatique et aurait conduit à l'existence du groupe des *Cyclobalanopsis*, puis du groupe asiatique *Ilex* (qui aurait migré en Europe) et d'un groupe qui aurait une répartition transcontinentale (hypothèse 1). La séparation des continents eurasiatique et américain aurait conduit au groupe des *Cerris* en Europe (qui aurait migrés ensuite vers l'Asie), et aux *Lobatae* d'un côté et des *Protobalanus* et des *Quercus s.s* d'un autre côté (ces derniers auraient par la suite migré en Asie via le détroit de Bering puis en Europe). Ou bien (hypothèse 2), en allant en Amérique, l'espèce ancestrale sœur des *Cerris* Européen aurait donné les *Lobatae*, puis les *Protobalanus* et les *Quercus s.s*.

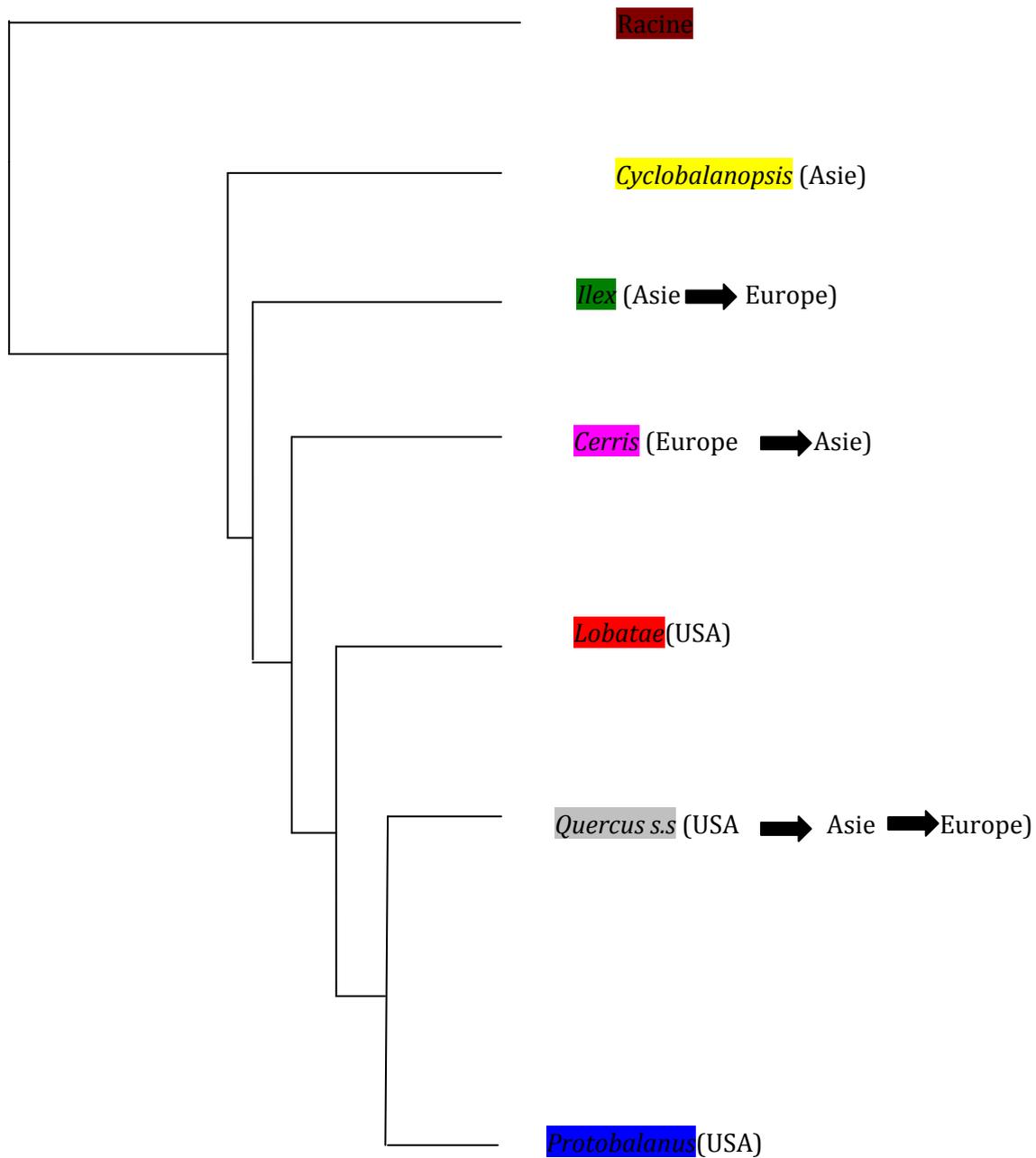


FIGURE B22 : PHYLOGENIE HYPOTHETIQUE BASEE SUR LES DONNEES ISSUES DE LA LITTERATURE (MANOS *ET AL.*, 1999; MANOS AND STANFORD, 2001; MANOS *ET AL.*, 2001; OH AND MANOS, 2008; DENK AND GRIMM, 2010) AINSI QUE SUR LES PHYLOGENIES NUCLEAIRES ET CHLOROPLASTIQUES DE LA PRESENTE ETUDE. LES FLECHES INDIQUENT LE SENS PROBABLE DE MIGRATION.

## B6 CONCLUSION GENERALE - PERSPECTIVES

---

La motivation initiale de ce travail était de mieux résoudre la phylogénie du genre *Quercus* grâce à une exploration et exploitation des ressources génomiques disponibles, et par la prise en compte d'un plus grand nombre d'espèces. Cette motivation provenait de l'irrésolution au niveau des nœuds les plus profonds des phylogénies réalisées avec des marqueurs nucléaires traditionnels (Manos *et al.*, 1999; Manos and Stanford, 2001; Manos *et al.*, 2001; Oh and Manos, 2008; Denk and Grimm, 2010), et caractérisées par cette séparation en deux groupes «New World Oaks» et «Old World Oaks». La raison biologique de cette irrésolution pouvait venir du fait que les spéciations ont été très rapides dans le passé et, qu'ainsi, les marqueurs moléculaires n'ont pas eu le temps de capter ce signal, et/ou que ceux qui ont pu le faire ont perdu l'information suite à la saturation des sites. Ce phénomène a donc atteint les différents types de génomes (nucléaires et chloroplastiques) et cette empreinte doit se retrouver dans chacun de ces types de marqueurs.

Afin de résoudre cette irrésolution, un échantillonnage plus important (112 espèces pour le génome nucléaire et 57 pour le génome chloroplastique) a été réalisé, en même temps que la recherche de nouveaux marqueurs nucléaires (parties B2 et B3) et le séquençage du génome chloroplastique entier sur un sous-échantillon (partie B4).

Au-delà de la résolution des nœuds les plus profonds grâce à l'apport de marqueurs nucléaires nouveaux et grâce à l'accès au génome chloroplastique complet, cette étude avait également pour ambition :

- 1) de confirmer ou d'infirmer l'existence de six groupes taxonomiques (*Cyclobalanopsis*, *Cerris*, *Ilex*, *Protobalanus*, *Lobatae* et *Quercus s.s.*),
- 2) de tester si le groupe des *Cyclobalanopsis* pouvait être considéré comme un sous-genre du genre *Quercus*,
- 3) de voir si les groupes *Cerris* et *Ilex* étaient distincts et, si tel est le cas, de voir si l'un était antérieur à l'autre,
- 4) de stabiliser la place des chênes intermédiaires du groupe des *Protobalanus* vis-à-vis des groupes *Lobatae* et *Quercus s.s.*,



- 5) de trouver des calibrations fossiles grâce aux travaux bibliographiques réalisés au sein du laboratoire et de dater les phylogénies moléculaires obtenues,
- 6) de synthétiser les informations recueillies afin d'établir un scénario de biogéographie historique du genre *Quercus*.

#### Au niveau nucléaire :

Deux types de marqueurs ont été prospectés :

- quatre marqueurs nucléaires de types COS (Conserved Ortholog Set markers, partie B2) qui n'ont pas permis d'obtenir d'informations, leur orthologie étant remise en cause,
- six marqueurs nucléaires développés dans le cadre d'un projet de reséquençage mené dans le réseau d'excellence Evoltree. Ces marqueurs ont été bien plus informatifs, et ils font partie intégrante d'un manuscrit soumis pour publication (partie B3).

Les deux groupes « Old World Oaks » et « New World Oaks » ont été retrouvés après concaténation en une SuperMatrice des six marqueurs nucléaires. Les valeurs de soutien aux branches sont relativement faibles (Figures B10 et B11) mais les six groupes taxonomiques semblent confirmés et bien soutenus (Figures B10 et B11).

Il n'est pas possible de bien séparer le groupe des *Cyclobalanopsis* des autres à ce stade (Figures B10 et B11). Cependant, lorsque la phylogénie est racinée sur le genre *Castanea* (Figure B10), il y a formation d'une longue branche attirant le groupe des *Cyclobalanopsis*. Ceci indique que le groupe des *Cyclobalanopsis* apparaît bien différent génétiquement des autres groupes. Mais sa position versatile en fonction de l'outgroup choisi (*Castanea* ou *Notholithocarpus*), et le faible échantillonnage des espèces de *Cyclobalanopsis* ne nous permet pas de conclure.

Les groupes *Cerris* et *Ilex* semblent distincts, bien qu'une espèce d'*Ilex* asiatique (*Q. franchetii*, Figures B10 et B11) se retrouve au sein des *Cerris*.

Encore une fois, il est possible d'arguer que ceci soit un artéfact, dû à la faible représentation en nombre d'espèces d'*Ilex* asiatique.

Nos résultats montrent que les groupes *Protobalanus* et *Quercus s.s* sont associés en groupes frères comme suggéré par le fort soutien aux branches (Figures B10 et B11). Ceci avait déjà été démontré sur des données ITS (Manos *et al.*, 1999), alors que, des études postérieures hésitaient sur cette place (Manos and Stanford, 2001; Manos *et al.*, 2001; Oh and Manos, 2008; Denk and Grimm, 2010).



La recherche de datations fossiles pour calibrer chacun des six groupes a été effectuée avec succès. Malheureusement, des problèmes d'ordre technique sont apparus. Ceux-ci proviennent du manque de consistance de la phylogénie nucléaire (l'ajout de datations fossiles modifiant la topologie de l'arbre ainsi inféré), mais également des datations elles-mêmes qui méritent d'être testées une par une afin de voir si les *a priori* et les *a posteriori* correspondent.

La découverte de nouveaux marqueurs nucléaires doit être poursuivie afin de conforter encore plus l'hypothèse de radiation rapide. En effet, nous nous attendons à ce que l'irrésolution aux nœuds les plus profonds s'amenuise à mesure que de nouvelles séquences sont incluses dans l'analyse.

L'échantillonnage taxonomique doit être complété sur les groupes des *Cyclobalanopsis* et des *Ilex* asiatiques afin d'avoir une image certainement plus juste de la variabilité existante entre ces groupes.

Cet effort devrait permettre d'utiliser les calibrations fossiles et donc de dater les événements de spéciation de façon plus correcte qu'actuellement. Ces travaux devront se poursuivre par l'utilisation de modèles d'inférence de scénarios de biogéographie historique qui sera menée à la suite de cette thèse dans le cadre de la dernière année du projet « PhyloSPACE ». L'apport de cartes paléoclimatiques dans le cadre de ce même projet contribuera à asseoir les hypothèses de biogéographie dégagées.

#### Au niveau chloroplastique :

Le séquençage complet du génome chloroplastique (plus de 160 000 pb) sur 57 espèces a fourni bon nombre de réponses aux questions soulevées (partie B4).

Le fait que la phylogénie moléculaire basée sur l'intégralité du génome chloroplastique soit bien mieux résolue que sur seulement un tiers des données (Figures B18 et B20) indique sans équivoque que le signal phylogénétique est diffus dans le génome, et qu'un nombre important de nucléotides est nécessaire pour augmenter la résolution.

Le groupe des *Cyclobalanopsis* y est bien différencié (Figure B18). Le groupe des *Ilex* est séparé en deux groupes géographiques (Figure B19) : les *Ilex* asiatiques et les *Ilex* européens. Le fait que les *Ilex* asiatiques apparaît comme groupe frère des *Cyclobalanopsis*, peut, comme évoqué précédemment, provenir d'une faiblesse de l'échantillonnage taxonomique des *Cyclobalanopsis* et dans une moindre mesure, des *Ilex* asiatiques. Mais l'intérêt majeur de cette phylogénie provient du fait que le groupe des *Ilex* apparaît sans équivoque comme plus ancien que le groupe *Cerris* (Figure B18). Les *Quercus* s.s sont séparés en deux sous-groupes suivant



leur géographie (Figure B19) : un sous-groupe américain et un sous-groupe eurasiatique. Le signal géographique du génome chloroplastique est donc fort, mais il ne faut pas le confondre avec un signal phylogénétique. Autrement dit, la phylogénie basée sur le chloroplaste donne des informations sur la migration des différents groupes, la composition de ces groupes étant reflétée par la phylogénie nucléaire. Il n'est pas possible à parti du génome chloroplastique et de l'échantillonnage actuel de bien visualiser l'histoire des *Protobalanus*, *Quercus s.s* et *Lobatae*. En effet, un *Protobalanus* se trouve affecté au groupe des *Quercus s.s* et un autre au groupe des *Lobatae*. Ces placements inattendus sont sans doute le résultat de capture cytoplasmique ancienne générée par des hybridations interspécifiques.

Il n'est pas possible à ce stade de fournir une phylogénie calibrée. En effet, la phylogénie obtenue est, contrairement à celle issue des données nucléaires, bien résolue ; mais le temps nécessaire pour l'obtenir dépasse largement le cadre de cette étude (plus de 5 mois). De plus, il est nécessaire de tester les datations une par une afin de les valider.

Dans l'idéal, il faudrait arriver à obtenir le même échantillon taxonomique entre les phylogénies nucléaires et chloroplastiques afin de les confronter et d'en tirer des informations sur l'histoire évolutive de ce groupe, mais également de voir si des phénomènes d'hybridation (détectés lors de l'incongruence entre les données nucléaires et chloroplastiques) peuvent être mis en évidence.

Pour conclure, cette étude a permis de façon certaine, de confirmer la subdivision du genre *Quercus* en six groupes taxonomiques (*Cyclobalanopsis*, *Cerris*, *Ilex*, *Protobalanus*, *Quercus s.s* et *Lobatae*) mais aussi d'affecter une origine asiatique à tout le genre *Quercus* (Figure B22). Le groupe *Ilex* a migré d'Asie en Europe, il a donné naissance au groupe *Cerris* (Figure B22) en Europe (puis il a migré en Asie) et a donné naissance aux Etats-Unis aux *Lobatae*, et au groupe composé de *Protobalanus* et de *Quercus s.s*. Les *Quercus s.s* américains ont migré en retour en Eurasie. Il est nécessaire d'affiner ce scénario par des méthodes issues de la biogéographie historique, en complément de phylogénies calibrées et de données paléoclimatiques. Ceci préfigurant la suite du travail qui sera fait dans la dernière année du projet PhyloSPACE. Il pourrait être très instructif de regarder sur le génome nucléaire, les taux de spéciation menant à chacun des six groupes, dans le but de voir de façon quantitative si l'évolution s'est réalisée par paliers et non pas de façon lente et continue (évolution dite ponctuée).



En termes de bilan, cette thèse se singularise par plusieurs apports nouveaux à la phylogénie du genre *Quercus* :

- elle confirme le caractère diffus du signal phylogénétique et par voie de conséquence, elle suggère que des progrès peuvent être obtenus par l'adjonction de séquences nouvelles,
- par rapport à la classification, elle confirme la subdivision du genre en six groupes, dont les relations phylogénétiques ont été précisées, même si certaines irrésolutions persistent. Ces résultats mettent notamment en cause la classification de Camus généralement admise dans le monde francophone,
- elle met très clairement en évidence l'empreinte phylogéographique dans le génome chloroplastique au niveau du genre et de sa distribution mondiale,
- le signal phylogéographique chloroplastique ajouté à la phylogénie nucléaire permet d'échafauder un scénario biogéographique de diversification du genre

En termes de perspective, cette thèse appelle à une étape supplémentaire et ultime qu'il conviendrait de réaliser. Sur la lancée de la recherche de gènes nucléaires entreprise dans cette thèse, elle consisterait à explorer de manière plus exhaustive encore le génome jusqu'à saturation du signal phylogénétique. Une telle approche deviendra techniquement possible grâce au séquençage global du génome nucléaire actuellement en cours. Cette étape devra également considérer d'étendre l'échantillonnage à toutes les espèces du genre. Un tel projet est actuellement en cours en Amérique du Nord et financé par la NSF. Il serait souhaitable de le compléter par une approche similaire en Eurasie.



## B6.1 RÉFÉRENCES BIBLIOGRAPHIQUES

---

- Denk, T. and G. W. Grimm (2010). "The oaks of western Eurasia: Traditional classifications and evidence from two nuclear markers." Taxon **59**(2): 351-366.
- Manos, P. S., *et al.* (1999). "Phylogeny, Biogeography, and Processes of Molecular Differentiation in *Quercus* Subgenus *Quercus* (Fagaceae)." Molecular Phylogenetics and Evolution **12**(3): 333-349.
- Manos, P. S. and A. M. Stanford (2001). "The Historical Biogeography of Fagaceae: Tracking the Tertiary History of Temperate and Subtropical Forests of the Northern Hemisphere." International Journal of Plant Sciences **162**(S6): S77-S93.
- Manos, P. S., *et al.* (2001). "Systematics of Fagaceae: Phylogenetic Tests of Reproductive Trait Evolution." International Journal of Plant Sciences **162**(6): 1361-1379.
- Oh, S. H. and P. S. Manos (2008). "Molecular phylogenetics and cupule evolution in Fagaceae as inferred from nuclear CRABS CLAW sequences." Taxon **57**(2): 434-451.



## B7 : ANNEXES

---

ANNEXE 1 : ORIGINE GÉOGRAPHIQUE DES ESPÈCES ET SUCCÈS DES AMPLIFICATIONS DES GÈNES COS.

Section/Groupe	Espèce	Pays	Latitude	Longitude	COS6785	COS5799	COS9687	COS10808
<i>Castanea</i>	<i>Castanea mollissima</i>	Chine			1	1	1	1
<i>Castanea</i>	<i>Castanea mollissima</i>	Chine			1	1	1	1
<i>Cyclobalanopsis</i>	<i>Quercus vestita</i>	Thaïlande	18°48'N	98°54'E				
<i>Cyclobalanopsis</i>	<i>Cyclobalanopsis stenophylloides</i>	Japon			1		1	1
<i>Cyclobalanopsis</i>	<i>Quercus semiserrata</i>	Thaïlande	18°48'N	98°54'E				
<i>Cyclobalanopsis</i>	<i>Quercus morii</i>	Japon			1		1	1
<i>Cyclobalanopsis</i>	<i>Quercus salicina</i>	Japon	34°45' N	133°45'E	1			1
<i>Cyclobalanopsis</i>	<i>Cyclobalanopsis glaucoides</i>	Chine	25°00'N	102°02'E	1	1	1	1
<i>Cyclobalanopsis</i>	<i>Quercus argyrotricha</i>	Chine			1		1	1
<i>Cyclobalanopsis</i>	<i>Quercus multinervis</i>	Chine			1	1	1	1
<i>Cyclobalanopsis</i>	<i>Cyclobalanopsis augustunii</i>	Chine	25°00'N	102°02'E	1	1	1	1
<i>Cerris</i>	<i>Quercus suber</i>	France	44°38'N	1°10'W	1		1	1
<i>Cerris</i>	<i>Quercus libanii</i>	Iran	36°38'N	45°08'E	1	1	1	1

<i>Cerris</i>	<i>Quercus chenii</i>	Chine	32°04'N	118°50'E	1		1	1
<i>Cerris</i>	<i>Quercus trojana</i>	Italie	40°37'N	17°10'E	1	1	1	1
<i>Cerris</i>	<i>Quercus castaneifolia</i>	Iran	36°34'N	52°03'E	1	1	1	1
<i>Cerris</i>	<i>Quercus brantii</i>	Israël	32°44'N	35°11'E	1	1	1	1
<i>Cerris</i>	<i>Quercus variabilis</i>	Chine	32°04'N	118°50'E	1		1	1
<i>Cerris</i>	<i>Quercus afares</i>	Algérie	36°46'N	3°40'E	1	1	1	1
<i>Cerris</i>	<i>Quercus ithaburensis</i>	Turquie	38°58'N	29°24'E	1	1	1	1
<i>Cerris</i>	<i>Quercus acutissima</i>	Chine	32°04'N	118°05'E	1	1	1	1
<i>Lobatae</i>	<i>Quercus castanea</i>	Mexique	20° 38'N	98° 36'W		1	1	1
<i>Lobatae</i>	<i>Quercus shumardii</i>	USA (Arkansas)	33°05'N	93°30'W	1		1	
<i>Lobatae</i>	<i>Quercus sideroxyla</i>	Turquie	19° 45'N	101° 15'W	1			1
<i>Lobatae</i>	<i>Quercus benthamii</i>	Mexique	17°05'N	93°2'W	1	1	1	1
<i>Lobatae</i>	<i>Quercus konzattii</i>	Mexique	17° 05'N	96° 33'W	1		1	1
<i>Lobatae</i>	<i>Quercus costaricensis</i>	Costa Rica				1	1	1
<i>Lobatae</i>	<i>Quercus crispipilis</i>	Mexique	16°47'N	92°5'W	1	1	1	1
<i>Lobatae</i>	<i>Quercus falcata</i>	USA (Louisiane)	32°30'N	92°40'W	1	1	1	1
<i>Lobatae</i>	<i>Quercus laurina</i>	Mexique	17° 25'N	96° 35'W	1		1	1

<i>Lobatae</i>	<i>Quercus sapoteifolia</i>	Mexique	16°06'N	91°40'W	1	1	1	1
<i>Lobatae</i>	<i>Quercus wislizenii</i>	USA (Californie)	33°42'N	117°25'W	1	1		1
<i>Lobatae</i>	<i>Quercus affinis</i>	Mexique	20° 40'N	98° 38'W		1	1	1
<i>Lobatae</i>	<i>Quercus candicans</i>	Mexique	16°06'N	91°40'W	1	1	1	1
<i>Lobatae</i>	<i>Quercus georgiana</i>	USA			1	1	1	1
<i>Lobatae</i>	<i>Quercus hypoleucoides</i>	USA (Nouveau Mexique)			1	1		1
<i>Lobatae</i>	<i>Quercus pinnativenulosa</i>	Mexique			1	1	1	1
<i>Lobatae</i>	<i>Quercus rysophylla</i>	Mexique			1		1	
<i>Lobatae</i>	<i>Quercus sartorii</i>	Mexique	20° 38'N	98° 36'W	1		1	1
<i>Lobatae</i>	<i>Quercus acherdophylla</i>	Mexique	20°14'N	98°12'W	1	1		1
<i>Lobatae</i>	<i>Quercus crassipes</i>	Mexique	19°35'N	102°05'W	1		1	1
<i>Lobatae</i>	<i>Quercus depressa</i>	Mexique	20°05'N	98°33'W	1	1	1	1
<i>Lobatae</i>	<i>Quercus hirtifolia</i>	Mexique	19°50'N	97°35'W	1	1	1	1
<i>Lobatae</i>	<i>Quercus myrtifolia</i>	USA (Floride)			1	1	1	1
<i>Lobatae</i>	<i>Quercus rubra</i>	USA (Kentucky)	37°10'N	88°20'W	1	1	1	1
<i>Lobatae</i>	<i>Quercus agrifolia</i>	USA (Californie)	33°42'N	117°25'W	1	1	1	1
<i>Lobatae</i>	<i>Quercus eugenifolia</i>	Mexique	20°10'N	98°14'W	1	1	1	1

<i>Lobatae</i>	<i>Quercus kelloggii</i>	USA (Californie)	37°36'N	120°02'W	1	1	1	1
<i>Lobatae</i>	<i>Quercus mexicana</i>	Mexique	20°07'N	98°18'W	1	1	1	1
<i>Lobatae</i>	<i>Quercus phellos</i>	USA (Louisianne)	32°30'N	92°40'W	1	1	1	1
<i>Lobatae</i>	<i>Quercus dysophylla</i>	Mexique	19°44'N	98°03'W	1	1	1	1
<i>Lobatae</i>	<i>Quercus palustris</i>	USA (Vermont)	41°35'N	72°20'W	1	1	1	1
<i>Ilex</i>	<i>Quercus dolicholepis</i>	Chine			1		1	1
<i>Ilex</i>	<i>Quercus franchetii</i>	Chine	25°05'N	102°26'E	1	1	1	1
<i>Ilex</i>	<i>Quercus coccifera</i>	Turquie	37°00'N	30°30'E	1	1	1	1
<i>Ilex</i>	<i>Quercus calliprinos</i>	Israël	32° 40'N	35°05'E	1	1	1	1
<i>Ilex</i>	<i>Quercus ilex</i>	France	44°38'N	1°10'W	1	1	1	1
<i>Quercus s.s</i>	<i>Quercus faginea</i>	Espagne	42°34'N	0°34'W	1			1
<i>Quercus s.s</i>	<i>Quercus lobata</i>	USA (Californie)	34°27'N	118°20'W	1	1		1
<i>Quercus s.s</i>	<i>Quercus malacotricha</i>	Chine	25°05'N	102°26'E	1		1	1
<i>Quercus s.s</i>	<i>Quercus toza</i>	France	44°44'N	0°45'W	1		1	1
<i>Quercus s.s</i>	<i>Quercus dentata</i>	Chine	32°04'N	118°50'E	1	1	1	1
<i>Quercus s.s</i>	<i>Quercus excelsa</i>	Mexique	20° 38'N	98° 36'W	1	1	1	1
<i>Quercus s.s</i>	<i>Quercus glandulifera</i>	Chine	32°04'N	118°50'E	1		1	1

<i>Quercus s.s</i>	<i>Quercus laeta</i>	Mexique	17° 25'N	96° 35'W	1		1	1
<i>Quercus s.s</i>	<i>Quercus macrocarpa</i>	USA (Kansas)	38°03'N	95°10'W	1	1	1	
<i>Quercus s.s</i>	<i>Quercus peduncularis</i>	Mexique	16°06'N	91°40'W	1	1	1	1
<i>Quercus s.s</i>	<i>Quercus pubescens</i>	France	45°18'N	0°54'W	1			1
<i>Quercus s.s</i>	<i>Quercus oleoides</i>	Cuba	22°37'N	83°39'W	1	1	1	1
<i>Quercus s.s</i>	<i>Quercus alba</i>	USA (Arkansas)	33°05'N	93°30'W	1	1	1	1
<i>Quercus s.s</i>	<i>Quercus berberidifolia</i>	USA (Californie)	33°18'N	116°38'W	1		1	1
<i>Quercus s.s</i>	<i>Quercus canariensis</i>	Algérie	36°47'N	5°46'E	1		1	1
<i>Quercus s.s</i>	<i>Quercus cornelius mulleri</i>	USA (Californie)	33°12'N	116°30'W	1		1	1
<i>Quercus s.s</i>	<i>Quercus germana</i>	Mexique			1	1		1
<i>Quercus s.s</i>	<i>Quercus glabrescens</i>	Mexique	20°07'N	98°37'W	1	1	1	1
<i>Quercus s.s</i>	<i>Quercus leiophylla</i>	Mexique	20° 38'N	98° 36'W	1	1	1	1
<i>Quercus s.s</i>	<i>Quercus muehlenbergii</i>	USA (Mississippi)	33°20'N	88°47'W	1	1		1
<i>Quercus s.s</i>	<i>Quercus prinus</i>	USA (New York)	43°16'N	77°37'W	1	1	1	
<i>Quercus s.s</i>	<i>Quercus pungens</i>	USA (Texas)			1	1	1	1
<i>Quercus s.s</i>	<i>Quercus sebifera</i>	Mexique	16°05'N	92°00'W	1	1	1	1
<i>Quercus s.s</i>	<i>Quercus segoviensis</i>	Mexique	16°47'N	92°05'W	1	1	1	1

<i>Quercus s.s</i>	<i>Quercus stellata</i>	USA (Texas)	30°19'N	94°41'W	1	1	1	1
<i>Quercus s.s</i>	<i>Quercus ajoensis</i>	USA (Arizona)			1	1	1	1
<i>Quercus s.s</i>	<i>Quercus chapmanii</i>	USA (Georgie)			1	1	1	1
<i>Quercus s.s</i>	<i>Quercus copeyensis</i>	Costa Rica			1	1	1	1
<i>Quercus s.s</i>	<i>Quercus durata</i>	USA (Californie)			1	1	1	1
<i>Quercus s.s</i>	<i>Quercus infectoria</i>	Iran	36°38'N	45°08'E	1		1	1
<i>Quercus s.s</i>	<i>Quercus pontica</i>	Turquie			1	1	1	1
<i>Quercus s.s</i>	<i>Quercus rugosa</i>	Mexique	17° 25'N	99° 35'W	1	1	1	1
<i>Quercus s.s</i>	<i>Quercus similis</i>	USA (Caroline du Sud)			1	1	1	1
<i>Quercus s.s</i>	<i>Quercus aliena</i>	Chine	32°04'N	118°50'E	1	1	1	1
<i>Quercus s.s</i>	<i>Quercus arizonica</i>	USA (Texas)			1	1	1	1
<i>Quercus s.s</i>	<i>Quercus boissieri</i>	Israël	32° 40'N	35°05'E	1	1	1	1
<i>Quercus s.s</i>	<i>Quercus corrugata</i>	Mexique	20°00'N	97°22'W	1	1	1	1
<i>Quercus s.s</i>	<i>Quercus dumosa</i>	USA (Californie)	33°42'N	117°25'W	1	1	1	1
<i>Quercus s.s</i>	<i>Quercus geminata</i>	USA (Géorgie)			1	1	1	1
<i>Quercus s.s</i>	<i>Quercus gregii</i>	Mexique	19°44'N	98°03'W	1	1	1	1
<i>Quercus s.s</i>	<i>Quercus pacifica</i>	Ile de Santa Cruz	33°59' N	119°43'W	1	1	1	1

<i>Quercus s.s</i>	<i>Quercus douglasii</i>	USA (Californie)	38°43'N	120°48'W	1	1	1	1
<i>Quercus s.s</i>	<i>Quercus garryana</i>	USA (Californie)	44°59'N	123°02'W	1	1	1	1
<i>Quercus s.s</i>	<i>Quercus griffithii</i>	Chine			1	1	1	1
<i>Quercus s.s</i>	<i>Quercus insignis</i>	Mexique	19° 55'N	96°58'W	1	1	1	1
<i>Quercus s.s</i>	<i>Quercus john-tuckeri</i>	USA	33°12'N	116°30'W	1	1	1	1
<i>Quercus s.s</i>	<i>Quercus petraea</i>	France	47°55'N	1°54'E	1	1	1	1
<i>Quercus s.s</i>	<i>Quercus robur</i>	France	44°44'N	0°42'W	1	1	1	1
<i>Quercus s.s</i>	<i>Quercus sadleriana</i>	USA (Californie)	41°59'N	123°31'W	1	1	1	1
<i>Protobalanus</i>	<i>Quercus palmeri</i>	USA (Californie)			1		1	1
<i>Protobalanus</i>	<i>Quercus tomentela</i>	Ile de Santa Cruz			1		1	1
<i>Protobalanus</i>	<i>Quercus chrysolepis</i>	USA (Californie)	37°10'N	121°25'W	1	1	1	1
<i>Protobalanus</i>	<i>Quercus vaccinifolia</i>	USA (Californie)			1	1	1	1

ANNEXE 2: GEOGRAPHIC ORIGIN, BOTANICAL GROUP BELONGINGS, NUMBER OF SPECIMENS AND AMPLIFIED GENES FOR ALL THE STUDIED SPECIES.

Species	Botanical group	Country (State)	Latitude	Longitude	Number of sequences	Gene 1155	Gene 8461	Gene 8561	Gene 5191	Gene 9715	Gene 8745
<i>Castanea mollissima</i>		China	NO	NO	1	1	1			1	
<i>Castanea mollissima</i>		China	NO	NO	1	1	1	1	1	1	
<i>Lithocarpus densiflorus var densiflorus</i>		USA (California)	41°58'N	123°30'W	1	1	1		1	1	
<i>Lithocarpus densiflorus var echinoides</i>		USA (California)	41°58'N	123°30'W	1	1		1	1	1	
<i>Quercus vestita</i>	<i>Cyclobalanopsis</i>	Thailand	18°48'N	98°54'E	2	1	1	1	1		1
<i>Quercus stenophylloides</i>	<i>Cyclobalanopsis</i>	Japan	NO	NO	1		1			1	
<i>Quercus semiserrata</i>	<i>Cyclobalanopsis</i>	Thailand	18°48'N	98°54'E	2	1	1	1	1	1	1
<i>Quercus morii</i>	<i>Cyclobalanopsis</i>	Japan	NO	NO	1		1			1	1
<i>Quercus salicina</i>	<i>Cyclobalanopsis</i>	Japan	34°45' N	133°45'E	1		1	1	1	1	1
<i>Quercus schottkyana</i>	<i>Cyclobalanopsis</i>	China	25°00'N	102°02'E	1	1		1	1		1
<i>Quercus argyrotricha</i>	<i>Cyclobalanopsis</i>	China	NO	NO	1	1	1		1	1	1
<i>Quercus multinervis</i>	<i>Cyclobalanopsis</i>	China	NO	NO	1		1	1		1	1
<i>Quercus augustunii</i>	<i>Cyclobalanopsis</i>	China	25°00'N	102°02'E	1	1	1		1	1	1
<i>Quercus suber</i>	<i>Cerris</i>	France	44°38'N	1°10'W	1			1	1	1	
<i>Quercus libanii</i>	<i>Cerris</i>	Iran	36°38'N	45°08'E	7	1	1			1	
<i>Quercus chenii</i>	<i>Cerris</i>	China	32°04'N	118°50'E	2	1		1		1	1
<i>Quercus trojana</i>	<i>Cerris</i>	Italy	40°37'N	17°10'E	2	1	1	1	1		
<i>Quercus castaneifolia</i>	<i>Cerris</i>	Iran	36°34'N	52°03'E	6	1	1		1	1	
<i>Quercus brantii</i>	<i>Cerris</i>	Israel	32°44'N	35°11'E	4	1		1		1	1
<i>Quercus variabilis</i>	<i>Cerris</i>	China	32°04'N	118°50'E	2	1	1	1		1	1
<i>Quercus afares</i>	<i>Cerris</i>	Algeria	36°46'N	3°40'E	3	1	1	1	1	1	
<i>Quercus ithaburensis</i>	<i>Cerris</i>	Turkey	38°58'N	29°24'E	11	1	1	1	1	1	1
<i>Quercus acutissima</i>	<i>Cerris</i>	China	32°04'N	118°05'E	2	1	1	1	1	1	1

<i>Quercus castanea</i>	<i>Lobatae</i>	Mexico	20° 38'N	98° 36'W	1			1		1
<i>Quercus shumardii</i>	<i>Lobatae</i>	USA (Arkansas)	33°05'N	93°30'W	1	1		1		1
<i>Quercus sideroxyla</i>	<i>Lobatae</i>	Mexico	19° 45'N	101° 15'W	2			1		1
<i>Quercus benthamii</i>	<i>Lobatae</i>	Mexico	17°05'N	93°2'W	2			1		1
<i>Quercus conzattii</i>	<i>Lobatae</i>	Mexico	17° 05'N	96° 33'W	3			1		1
<i>Quercus costaricensis</i>	<i>Lobatae</i>	Costa Rica	NO	NO	5			1		1
<i>Quercus crispipilis</i>	<i>Lobatae</i>	Mexico	16°47'N	92°5'W	4			1		1
<i>Quercus falcata</i>	<i>Lobatae</i>	USA (Louisiana)	32°30'N	92°40'W	1	1				1
<i>Quercus laurina</i>	<i>Lobatae</i>	Mexico	17° 25'N	96° 35'W	4			1		1
<i>Quercus sapoteifolia</i>	<i>Lobatae</i>	Mexico	16°06'N	91°40'W	2			1		1
<i>Quercus wislizenii</i>	<i>Lobatae</i>	USA (California)	33°42'N	117°25'W	2	1		1		1
<i>Quercus affinis</i>	<i>Lobatae</i>	Mexico	20° 40'N	98° 38'W	2			1	1	1
<i>Quercus candicans</i>	<i>Lobatae</i>	Mexico	16°06'N	91°40'W	2			1		1
<i>Quercus georgiana</i>	<i>Lobatae</i>	No information USA (New	NO	NO	2	1	1			1
<i>Quercus hypoleuroides</i>	<i>Lobatae</i>	Mexico)	33°19'14"	108°50'19"	1	1		1		1
<i>Quercus pinnativenulosa</i>	<i>Lobatae</i>	Mexico	NO	NO	1	1	1			1
<i>Quercus rysophylla</i>	<i>Lobatae</i>	Mexico	25°36'55"	100°21'54"	1	1	1	1		1
<i>Quercus sartorii</i>	<i>Lobatae</i>	Mexico	20° 38'N	98° 36'W	2	1	1			1
<i>Quercus acherdophylla</i>	<i>Lobatae</i>	Mexico	20°14'N	98°12'W	2	1	1	1		1
<i>Quercus crassipes</i>	<i>Lobatae</i>	Mexico	19°35'N	102°05'W	1	1	1	1		1
<i>Quercus depressa</i>	<i>Lobatae</i>	Mexico	20°05'N	98°33'W	1	1	1	1		1
<i>Quercus hirtifolia</i>	<i>Lobatae</i>	Mexico	19°50'N	97°35'W	1	1	1	1		1
<i>Quercus myrtifolia</i>	<i>Lobatae</i>	USA (Florida)	29°53'15"	84°22'19"	2	1	1	1		1
<i>Quercus rubra</i>	<i>Lobatae</i>	USA (Kentucky)	37°10'N	88°20'W	1	1			1	1
<i>Quercus agrifolia</i>	<i>Lobatae</i>	USA (California)	33°42'N	117°25'W	3	1	1	1		1
<i>Quercus eugenifolia</i>	<i>Lobatae</i>	Mexico	20°10'N	98°14'W	3	1	1	1		1
<i>Quercus kelloggii</i>	<i>Lobatae</i>	USA (California)	37°36'N	120°02'W	3	1	1	1		1

<i>Quercus mexicana</i>	<i>Lobatae</i>	Mexico	20°07'N	98°18'W	2	1	1	1		1	1
<i>Quercus phellos</i>	<i>Lobatae</i>	USA (Louisiana)	32°30'N	92°40'W	3	1	1	1		1	1
<i>Quercus dysophylla</i>	<i>Lobatae</i>	Mexico	19°44'N	98°03'W	2	1	1	1	1	1	1
<i>Quercus palustris</i>	<i>Lobatae</i>	USA (Vermont)	41°35'N	72°20'W	2	1	1	1	1	1	1
<i>Quercus dolicholepis</i>	<i>Ilex</i>	China	NO	NO	1	1	1			1	
<i>Quercus franchetii</i>	<i>Ilex</i>	China	25°05'N	102°26'E	1			1		1	1
<i>Quercus coccifera</i>	<i>Ilex</i>	Turkey	37°00'N	30°30'E	5	1	1		1	1	1
<i>Quercus calliprinos</i>	<i>Ilex</i>	Israel	32° 40'N	35°05'E	10	1	1	1	1	1	1
<i>Quercus ilex</i>	<i>Ilex</i>	France	44°38'N	1°10'W	4	1	1	1	1	1	1
<i>Quercus faginea</i>	<i>Quercus s.s.</i>	Spain	42°34'N	0°34'W	1	1		1		1	
<i>Quercus lobata</i>	<i>Quercus s.s.</i>	USA (California)	34°27'N	118°20'W	1			1		1	
<i>Quercus malacotricha</i>	<i>Quercus s.s.</i>	China	25°05'N	102°26'E	1			1			1
<i>Quercus toza</i>	<i>Quercus s.s.</i>	France	44°44'N	0°45'W	1	1				1	
<i>Quercus dentata</i>	<i>Quercus s.s.</i>	China	32°04'N	118°50'E	1	1				1	
<i>Quercus excelsa</i>	<i>Quercus s.s.</i>	Mexico	20° 38'N	98° 36'W	2					1	1
<i>Quercus glandulifera</i>	<i>Quercus s.s.</i>	China	32°04'N	118°50'E	1	1		1	1		
<i>Quercus laeta</i>	<i>Quercus s.s.</i>	Mexico	17° 25'N	96° 35'W	2			1	1		1
<i>Quercus macrocarpa</i>	<i>Quercus s.s.</i>	USA (Kansas)	38°03'N	95°10'W	1	1				1	1
<i>Quercus peduncularis</i>	<i>Quercus s.s.</i>	Mexico	16°06'N	91°40'W	2			1	1		
<i>Quercus pubescens</i>	<i>Quercus s.s.</i>	France	45°18'N	0°54'W	1	1		1	1		1
<i>Quercus oleoides</i>	<i>Quercus s.s.</i>	Cuba	22°37'N	83°39'W	3			1		1	1
<i>Quercus alba</i>	<i>Quercus s.s.</i>	USA (Arkansas)	33°05'N	93°30'W	1	1			1	1	
<i>Quercus berberidifolia</i>	<i>Quercus s.s.</i>	USA (California)	33°18'N	116°38'W	1	1	1		1		1
<i>Quercus canariensis</i>	<i>Quercus s.s.</i>	Algeria	36°47'N	5°46'E	1	1		1		1	1
<i>Quercus cornelius mulleri</i>	<i>Quercus s.s.</i>	USA (California)	33°12'N	116°30'W	1	1	1			1	1
<i>Quercus germana</i>	<i>Quercus s.s.</i>	Mexico	NO	NO	1	1	1			1	1
<i>Quercus glabrescens</i>	<i>Quercus s.s.</i>	Mexico	20°07'N	98°37'W	2	1	1				1
<i>Quercus leiophylla</i>	<i>Quercus s.s.</i>	Mexico	20° 38'N	98° 36'W	1			1		1	1

<i>Quercus muehlenbergii</i>	<i>Quercus s.s.</i>	USA(Mississippi)	33°20'N	88°47'W	1	1		1		1	1
<i>Quercus montana</i>	<i>Quercus s.s.</i>	USA (New York)	43°16'N	77°37'W	1			1	1	1	1
<i>Quercus pungens</i>	<i>Quercus s.s.</i>	USA (Texas)	32°0'13"	104°47'19"	1		1	1		1	
<i>Quercus sebifera</i>	<i>Quercus s.s.</i>	Mexico	16°05'N	92°00'W	2			1	1		1
<i>Quercus segoviensis</i>	<i>Quercus s.s.</i>	Mexico	16°47'N	92°05'W	2			1		1	1
<i>Quercus stellata</i>	<i>Quercus s.s.</i>	USA (Texas)	30°19'N	94°41'W	1	1			1	1	
<i>Quercus ajoensis</i>	<i>Quercus s.s.</i>	USA (Arizona)	NO	NO	1	1	1	1		1	1
<i>Quercus chapmanii</i>	<i>Quercus s.s.</i>	USA (Georgia)	31°22'15"	82°10'05"	1	1	1			1	1
<i>Quercus copeyensis</i>	<i>Quercus s.s.</i>	Costa Rica	NO	NO	6			1	1	1	1
<i>Quercus durata</i>	<i>Quercus s.s.</i>	USA (California)	38°59'20"	123°05'07"	2	1	1		1		1
<i>Quercus infectoria</i>	<i>Quercus s.s.</i>	Iran	36°38'N	45°08'E	5	1	1	1		1	1
<i>Quercus pontica</i>	<i>Quercus s.s.</i>	No information	NO	NO	1	1			1	1	1
<i>Quercus rugosa</i>	<i>Quercus s.s.</i>	Mexico USA (South Carolina)	17° 25'N	99° 35'W	5			1	1	1	1
<i>Quercus similis</i>	<i>Quercus s.s.</i>	Carolina)	32°51'38"	79°54'08"	1	1	1			1	1
<i>Quercus aliena</i>	<i>Quercus s.s.</i>	China	32°04'N	118°50'E	2	1		1	1	1	1
<i>Quercus arizonica</i>	<i>Quercus s.s.</i>	USA (Texas)	31°59'17"	104°46'48"	1	1	1	1		1	1
<i>Quercus boissieri</i>	<i>Quercus s.s.</i>	Israel	32° 40'N	35°05'E	7	1	1	1		1	1
<i>Quercus corrugata</i>	<i>Quercus s.s.</i>	Mexico	20°00'N	97°22'W	1	1	1	1		1	1
<i>Quercus dumosa</i>	<i>Quercus s.s.</i>	USA (California)	33°42'N	117°25'W	1	1	1		1	1	1
<i>Quercus geminata</i>	<i>Quercus s.s.</i>	USA (Georgia)	32°0'32"	81°46'07"	2	1	1		1	1	1
<i>Quercus gregii</i>	<i>Quercus s.s.</i>	Mexico	19°44'N	98°03'W	1	1	1	1		1	1
<i>Quercus pacifica</i>	<i>Quercus s.s.</i>	USA (California )	33°59' N	119°43'W	2	1	1		1	1	1
<i>Quercus douglasii</i>	<i>Quercus s.s.</i>	USA (California)	38°43'N	120°48'W	3	1	1	1	1	1	1
<i>Quercus garryana</i>	<i>Quercus s.s.</i>	USA (California)	44°59'N	123°02'W	2	1	1	1	1	1	1
<i>Quercus griffithii</i>	<i>Quercus s.s.</i>	No information	NO	NO	1	1	1	1	1	1	1
<i>Quercus insignis</i>	<i>Quercus s.s.</i>	Mexico	19° 55'N	96°58'W	3	1	1	1	1	1	1
<i>Quercus john-tuckeri</i>	<i>Quercus s.s.</i>	USA	33°12'N	116°30'W	3	1	1	1	1	1	1

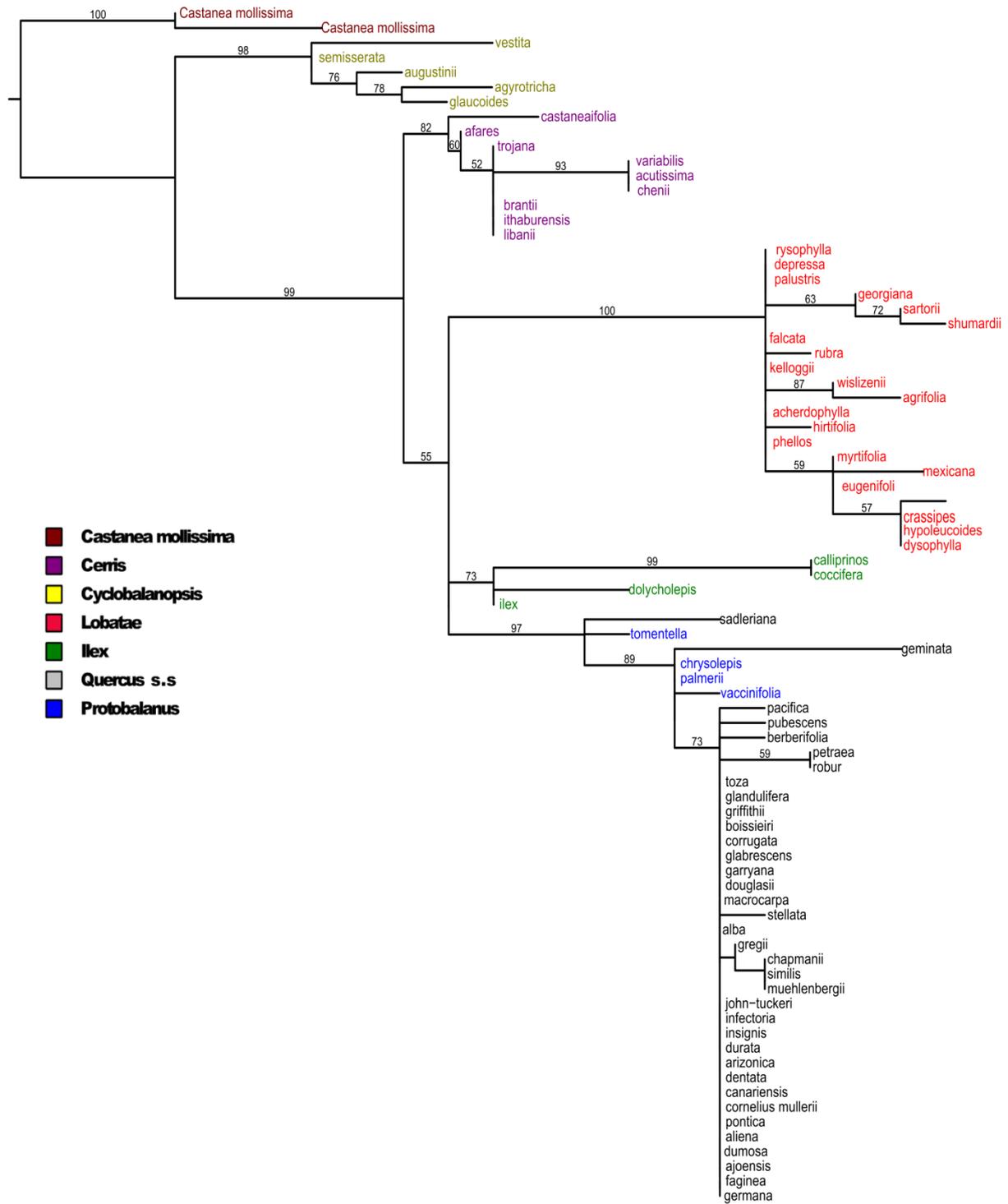
<i>Quercus petraea</i>	<i>Quercus s.s.</i>	France	47°55'N	1°54'E	3	1	1	1	1	1	1
<i>Quercus robur</i>	<i>Quercus s.s.</i>	France	44°44'N	0°42'W	2	1	1	1	1	1	1
<i>Quercus sadleriana</i>	<i>Quercus s.s.</i>	USA (California)	41°59'N	123°31'W	2	1	1	1	1	1	1
<i>Quercus palmeri</i>	<i>Protobalanus</i>	USA (California)	33°35'41"	116°36'32"	2	1				1	1
<i>Quercus tomentela</i>	<i>Protobalanus</i>	USA (California)	34°02'17"	119°36'34"	1	1	1	1			1
<i>Quercus chrysolepis</i>	<i>Protobalanus</i>	USA (California)	37°10'N	121°25'W	3	1		1	1	1	1
<i>Quercus vaccinifolia</i>	<i>Protobalanus</i>	USA (California)	39°43'08"	122°53'03"	2	1	1	1		1	1

ANNEXE 3: ACCESSION NUMBER OF THE COLLECTED SPECIES.

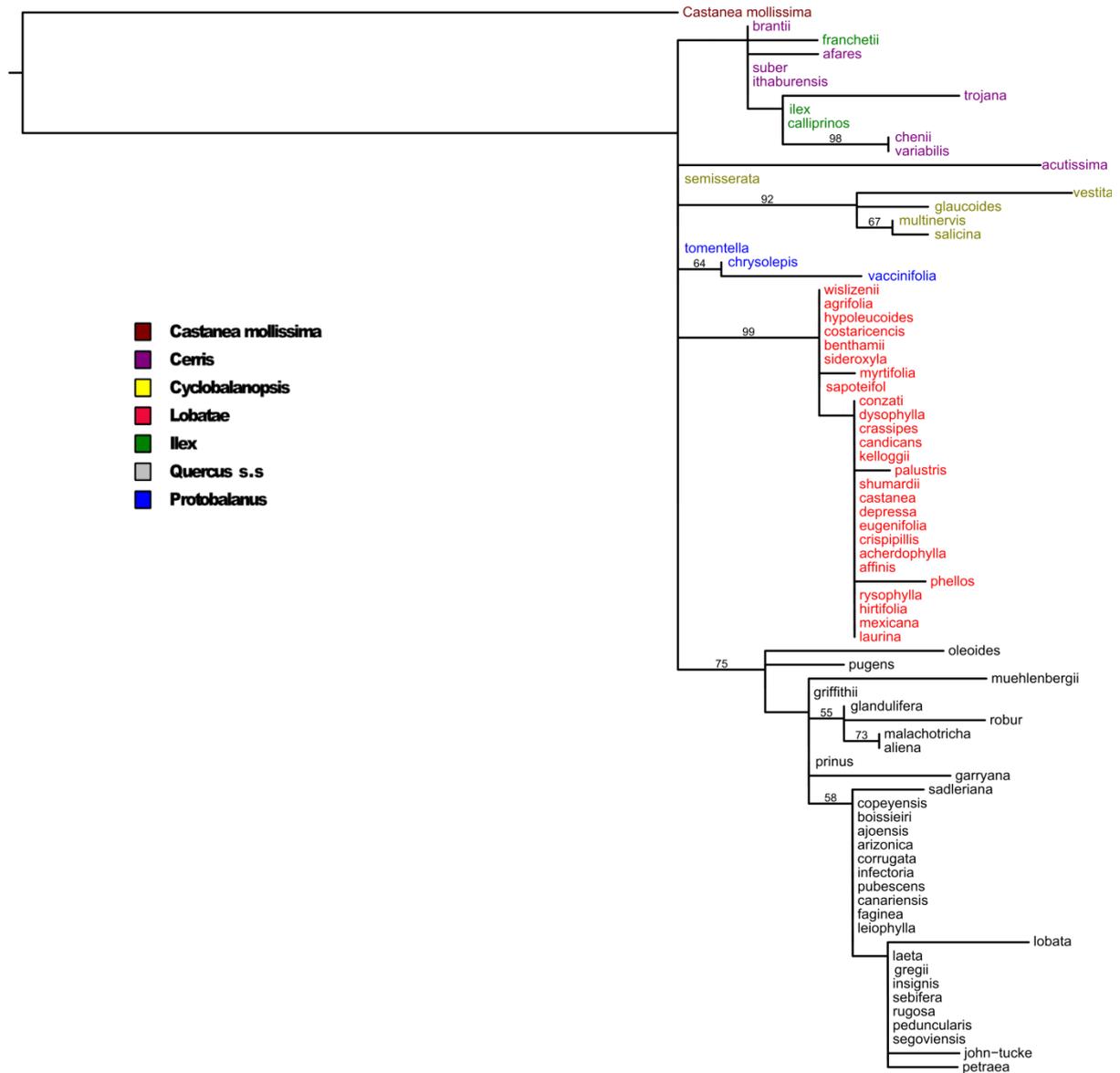
	1155	5191	8461	8561	8745	9715
<i>Lithocarpus densiflorus</i> <i>var echinoides</i>	HF680262	HF680264		HF680267		HF680268
<i>Castanea mollissima</i>	HF563134	HF563265	HF563326	HF563208		HF563494
<i>Castanea mollissima</i>	HF563133		HF563325			HF563493
<i>Quercus acherdophylla</i>	HF563080		HF563285	HF563161	HF563348	HF563434
<i>Quercus acutissima</i>	HF563081	HF563238	HF563286	HF563162	HF563349	HF563435
<i>Quercus afares</i>	HF563082	HF563239	HF563287	HF563163		HF563436
<i>Quercus affinis</i>		HF563240		HF563164	HF563350	HF563437
<i>Quercus agrifolia</i>	HF563083		HF563288	HF563165	HF563351	HF563438
<i>Quercus ajoensis</i>	HF563085		HF563290	HF563166	HF563353	HF563440
<i>Quercus alba</i>	HF563086	HF563242				HF563441
<i>Quercus aliena</i>	HF563087	HF563243		HF563167	HF563354	HF563442
<i>Quercus argyrotricha</i>	HF563084	HF563241	HF563289		HF563352	HF563439
<i>Quercus arizonica</i>	HF563088		HF563291	HF563168	HF563355	HF563443
<i>Quercus augustinii</i>	HF563089	HF563244	HF563292		HF563356	HF563444
<i>Quercus benthamii</i>				HF563169		HF563445
<i>Quercus berberidifolia</i>	HF563090	HF563245	HF563293		HF563357	
<i>Quercus brantii</i>	HF563092			HF563171	HF563359	HF563447
<i>Quercus canariensis</i>	HF563094			HF563173	HF563361	HF563449
<i>Quercus candicans</i>				HF563174	HF563362	HF563450
<i>Quercus castanea</i>				HF563175		HF563451
<i>Quercus castaneifolia</i>	HF563095	HF563247	HF563296			HF563452
<i>Quercus chapmanii</i>	HF563096		HF563297		HF563363	HF563453
<i>Quercus chenii</i>	HF563097			HF563176	HF563364	HF563454
<i>Quercus chrysolepis</i>	HF563098	HF563248		HF563177	HF563365	HF563455
<i>Quercus coccifera</i>	HF563099	HF563249	HF563298	HF563172	HF563366	HF563456
<i>Quercus konzattii</i>				HF563178	HF563367	HF563457
<i>Quercus copeyensis</i>		HF563250		HF563179	HF563368	HF563458
<i>Quercus cornelius</i>	HF563100		HF563299			HF563459
<i>Quercus corrugata</i>	HF563101		HF563300	HF563180	HF563369	HF563460
<i>Quercus costaricensis</i>				HF563181	HF563370	HF563461
<i>Quercus crassipes</i>	HF563102		HF563301	HF563182	HF563371	HF563462
<i>Quercus crispipilis</i>				HF563183	HF563372	
<i>Quercus dentata</i>	HF563103				HF563373	HF563463
<i>Quercus depressa</i>	HF563104		HF563302	HF563184	HF563374	
<i>Quercus dolicholepis</i>	HF563105		HF563303			HF563464
<i>Quercus douglasii</i>	HF563106	HF563251	HF563304		HF563375	HF563465
<i>Quercus dumosa</i>	HF563107	HF563252	HF563305		HF563376	HF563466

<i>Quercus durata</i>	HF563108	HF563253	HF563306		HF563377	
<i>Quercus dysophylla</i>	HF563109	HF563254	HF563307	HF563185	HF563378	HF563467
<i>Quercus eugeniifolia</i>	HF563110		HF563308	HF563186	HF563379	HF563468
<i>Quercus excelsa</i>					HF563380	HF563469
<i>Quercus faginea</i>	HF563111			HF563187		HF563470
<i>Quercus falcata</i>	HF563112					HF563471
<i>Quercus franchetii</i>				HF563188	HF563381	HF563472
<i>Quercus garryana</i>	HF563113	HF563255	HF563309	HF563189	HF563382	HF563473
<i>Quercus geminata</i>	HF563114	HF563256	HF563310		HF563383	HF563474
<i>Quercus georgiana</i>	HF563115		HF563311			HF563475
<i>Quercus germana</i>	HF563116		HF563312		HF563384	HF563476
<i>Quercus glabrescens</i>	HF563117		HF563313		HF563385	
<i>Quercus greggii</i>	HF563120		HF563314	HF563192	HF563387	HF563477
<i>Quercus griffithii</i>	HF563121	HF563259	HF563315	HF563193	HF563388	HF563478
<i>Quercus hirtifolia</i>	HF563122		HF563316	HF563194		HF563479
<i>Quercus hypoleucoides</i>	HF563123			HF563195	HF563389	HF563480
<i>Quercus ilex</i>	HF563124	HF563260	HF563317	HF563196	HF563390	HF563481
<i>Quercus infectoria</i>	HF563125		HF563318	HF563197	HF563391	HF563482
<i>Quercus insignis</i>	HF563126	HF563261	HF563319	HF563198	HF563392	HF563483
<i>Quercus ithaburensis</i>	HF563127	HF563262	HF563320	HF563199	HF563393	HF563484
<i>Quercus john</i>	HF563128	HF563263	HF563321	HF563200	HF563394	HF563485
<i>Quercus kelloggii</i>	HF563129		HF563322	HF563201	HF563395	HF563486
<i>Quercus laeta</i>		HF563264		HF563202	HF563396	
<i>Quercus laurina</i>				HF563203	HF563397	HF563487
<i>Quercus leiophylla</i>				HF563204	HF563398	HF563488
<i>Quercus libani</i>	HF563130		HF563323			HF563489
<i>Quercus lobata</i>				HF563205		HF563490
<i>Quercus macrocarpa</i>	HF563131				HF563399	HF563491
<i>Quercus malacotricha</i>				HF563206	HF563400	
<i>Quercus mexicana</i>	HF563132		HF563324	HF563207	HF563401	HF563492
<i>Quercus montana</i>		HF563271		HF563217	HF563414	HF563505
<i>Quercus morii</i>			HF563327		HF563402	HF563495
<i>Quercus muehlenbergii</i>	HF563135			HF563209	HF563403	HF563496
<i>Quercus multinervis</i>			HF563328	HF563210	HF563404	HF563497
<i>Quercus myrtifolia</i>	HF563136		HF563329	HF563211	HF563405	
<i>Quercus oleoides</i>				HF563212	HF563406	HF563498
<i>Quercus pacifica</i>	HF563137	HF563266	HF563330		HF563407	HF563499
<i>Quercus palmeri</i>	HF563138				HF563408	HF563500
<i>Quercus palustris</i>	HF563139	HF563267	HF563331	HF563213	HF563409	HF563501
<i>Quercus peduncularis</i>		HF563268		HF563214		
<i>Quercus petraea</i>	HF563140	HF563269	HF563332	HF563215	HF563410	HF563502

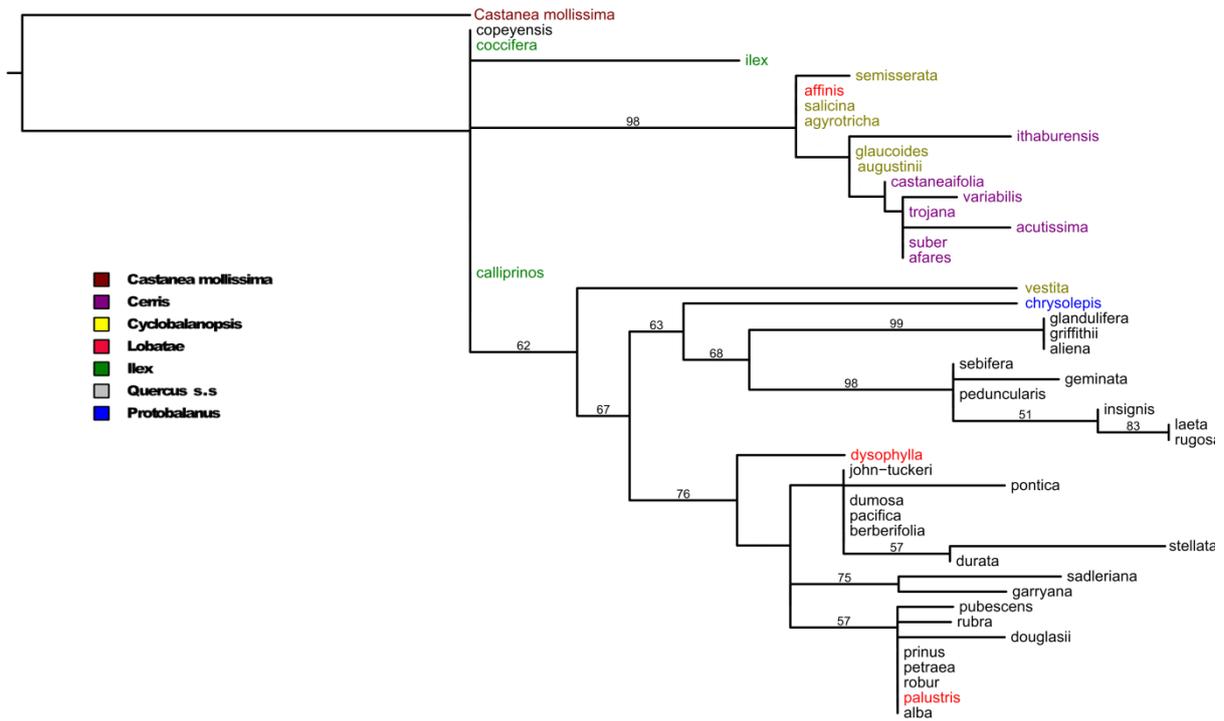
<i>Quercus phellos</i>	HF563141		HF563333	HF563216	HF563411	HF563503
<i>Quercus pinnativenulosa</i>	HF563142		HF563334		HF563412	
<i>Quercus pontica</i>	HF563143	HF563270			HF563413	HF563504
<i>Quercus pubescens</i>	HF563144	HF563272		HF563218	HF563415	
<i>Quercus pungens</i>			HF563335	HF563219		HF563506
<i>Quercus robur</i>	HF563145	HF563273	HF563336	HF563220	HF563416	HF563507
<i>Quercus rubra</i>	HF563146	HF563274			HF563417	HF563508
<i>Quercus rugosa</i>		HF563275		HF563221	HF563418	HF563509
<i>Quercus rhyssophylla</i>	HF563147		HF563337	HF563222	HF563419	HF563510
<i>Quercus sadleriana</i>	HF563148	HF563276	HF563338	HF563223	HF563420	HF563511
<i>Quercus salicina</i>		HF563277	HF563339	HF563224	HF563421	HF563512
<i>Quercus sapotifolia</i>				HF563225	HF563422	
<i>Quercus sartorii</i>	HF563149		HF563340		HF563423	HF563513
<i>Quercus schottkyana</i>	HF563119	HF563258		HF563191	HF563386	
<i>Quercus sebifera</i>		HF563278		HF563226	HF563424	
<i>Quercus segoviensis</i>				HF563227	HF563425	HF563514
<i>Quercus semiserrata</i>	HF563150	HF563279	HF563341	HF563228	HF563426	HF563515
<i>Quercus serrata</i>	HF563118	HF563257		HF563190		
<i>Quercus shumardii</i>	HF563151			HF563229	HF563427	
<i>Quercus sideroxyla</i>				HF563230	HF563428	HF563516
<i>Quercus similis</i>	HF563152		HF563342		HF563429	HF563517
<i>Quercus stellata</i>	HF563153	HF563280				HF563518
<i>Quercus stenophylloides</i>			HF563343			HF563519
<i>Quercus suber</i>		HF563281		HF563231		HF563520
<i>Quercus tomentella</i>	HF563154		HF563344	HF563232	HF563430	
<i>Quercus toza</i>	HF563155					HF563521
<i>Quercus trojana</i>	HF563156	HF563282	HF563345	HF563233		
<i>Quercus vacciniifolia</i>	HF563157		HF563346	HF563234	HF563431	HF563522
<i>Quercus variabilis</i>	HF563158	HF563283		HF563235	HF563432	HF563523
<i>Quercus vestita</i>	HF563159	HF563284	HF563347	HF563236	HF563433	
<i>Quercus wislizeni</i>	HF563160			HF563237		HF563524



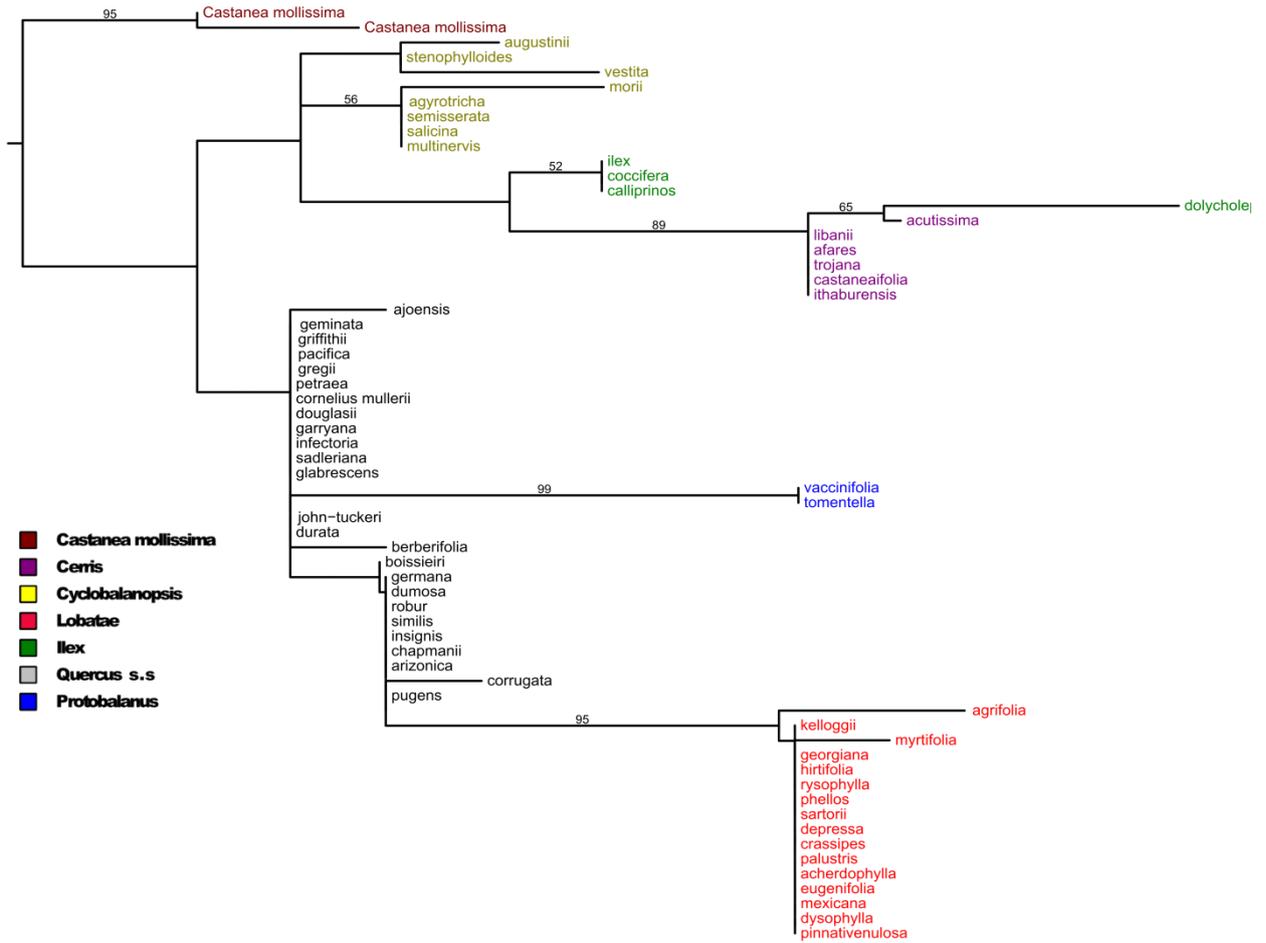
ANNEXE 4A: MAXIMUM LIKELIHOOD PHYLOGRAM OF THE GENUS QUERCUS BASED ON CL1155 AND ROOTED ON THE GENUS CASTANEA. NUMBERS CLOSE TO THE NODES REFER TO MAXIMUM LIKELIHOOD BOOTSTRAP SUPPORT.



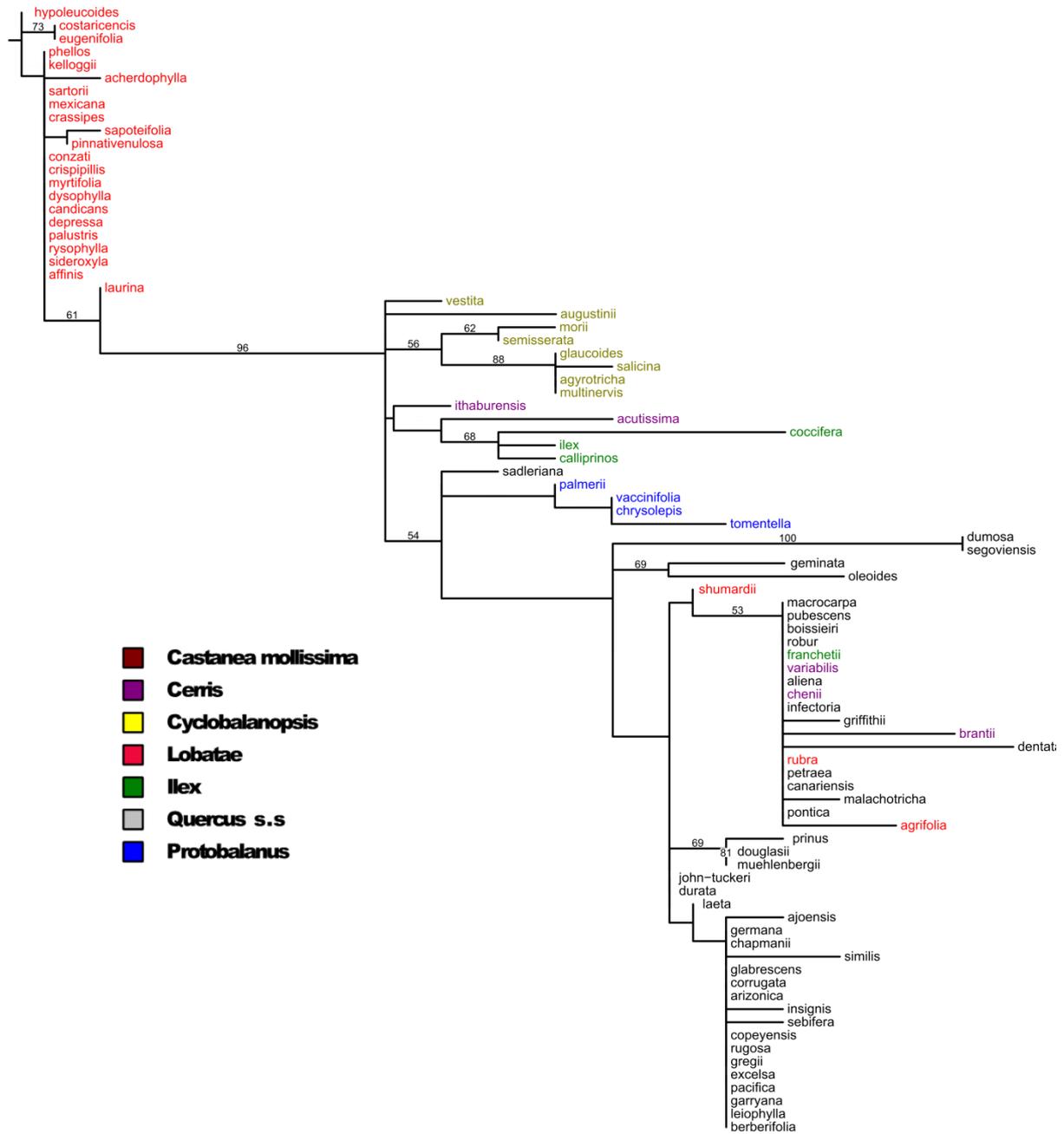
ANNEXE 4B: MAXIMUM LIKELIHOOD PHYLOGRAM OF THE GENUS QUERCUS BASED ON CL8561 AND ROOTED ON THE GENUS CASTANEA. NUMBERS CLOSE TO THE NODES REFER TO MAXIMUM LIKELIHOOD BOOTSTRAP SUPPORT.



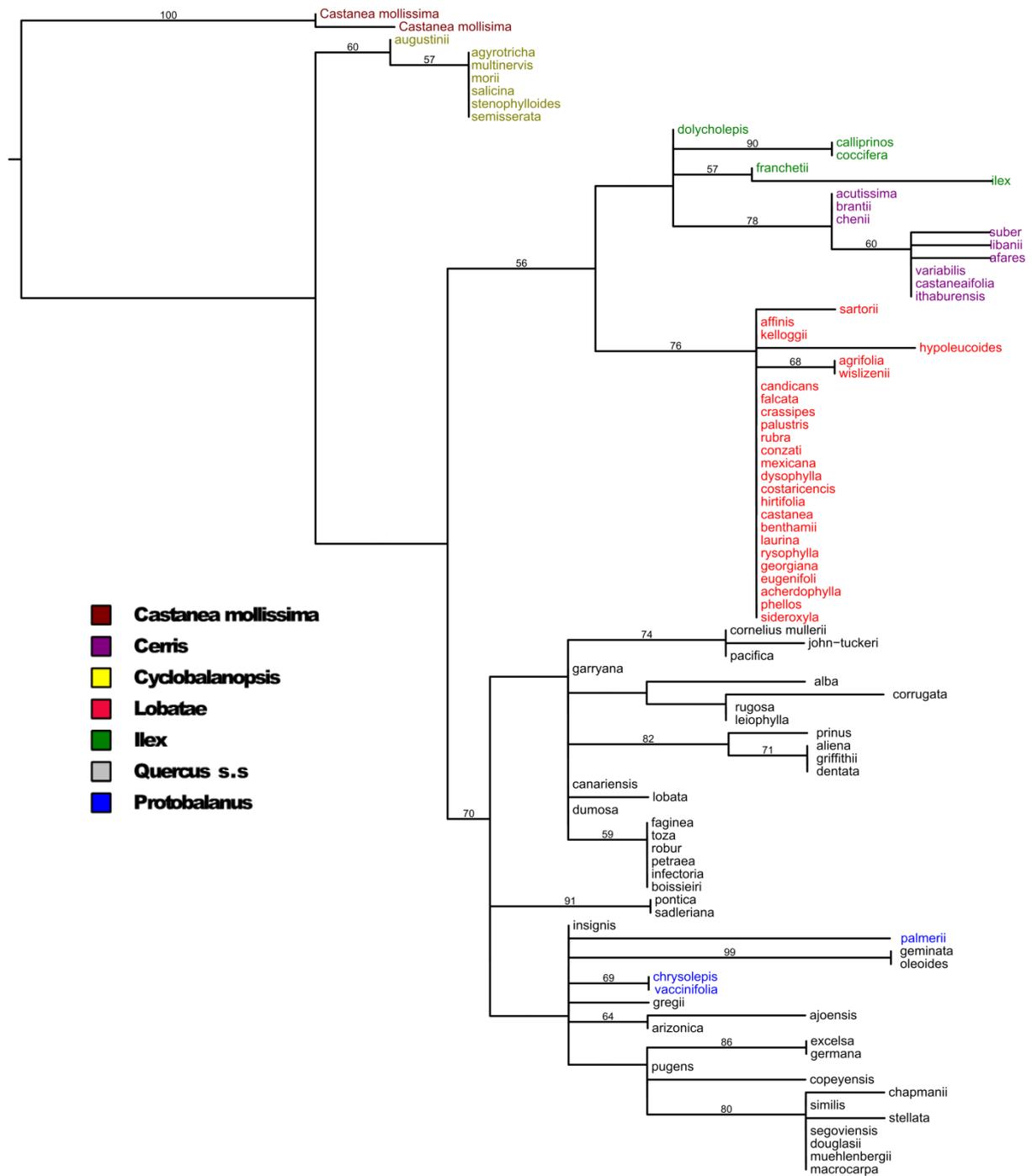
ANNEXE 4C: MAXIMUM LIKELIHOOD PHYLOGRAM OF THE GENUS QUERCUS BASED ON CL5191 AND ROOTED ON THE GENUS CASTANEA. NUMBERS CLOSE TO THE NODES REFER TO MAXIMUM LIKELIHOOD BOOTSTRAP SUPPORT.



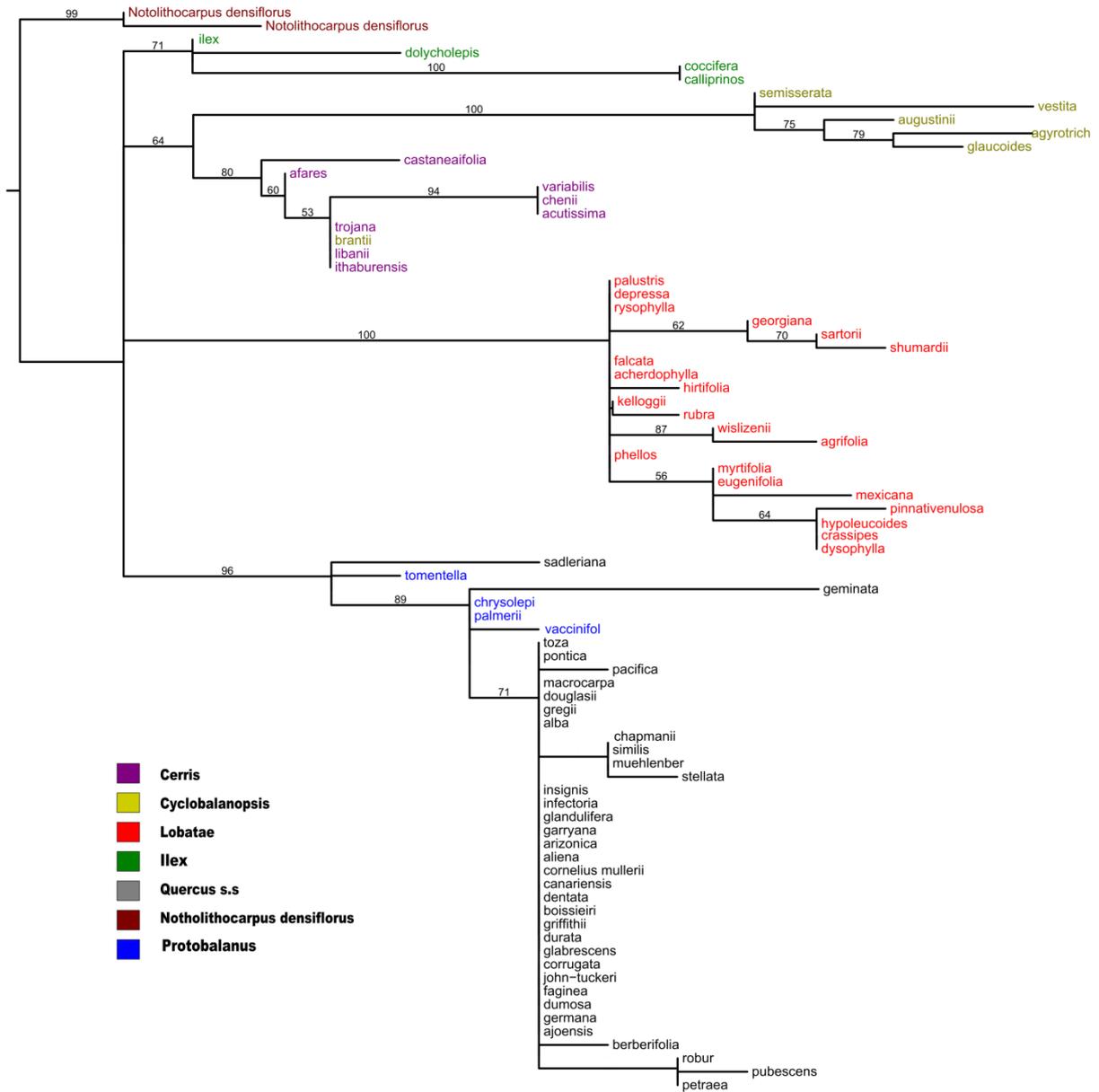
ANNEXE 4D: MAXIMUM LIKELIHOOD PHYLOGRAM OF THE GENUS QUERCUS BASED ON CL8461 AND ROOTED ON THE GENUS CASTANEA. NUMBERS CLOSE TO THE NODES REFER TO MAXIMUM LIKELIHOOD BOOTSTRAP SUPPORT.



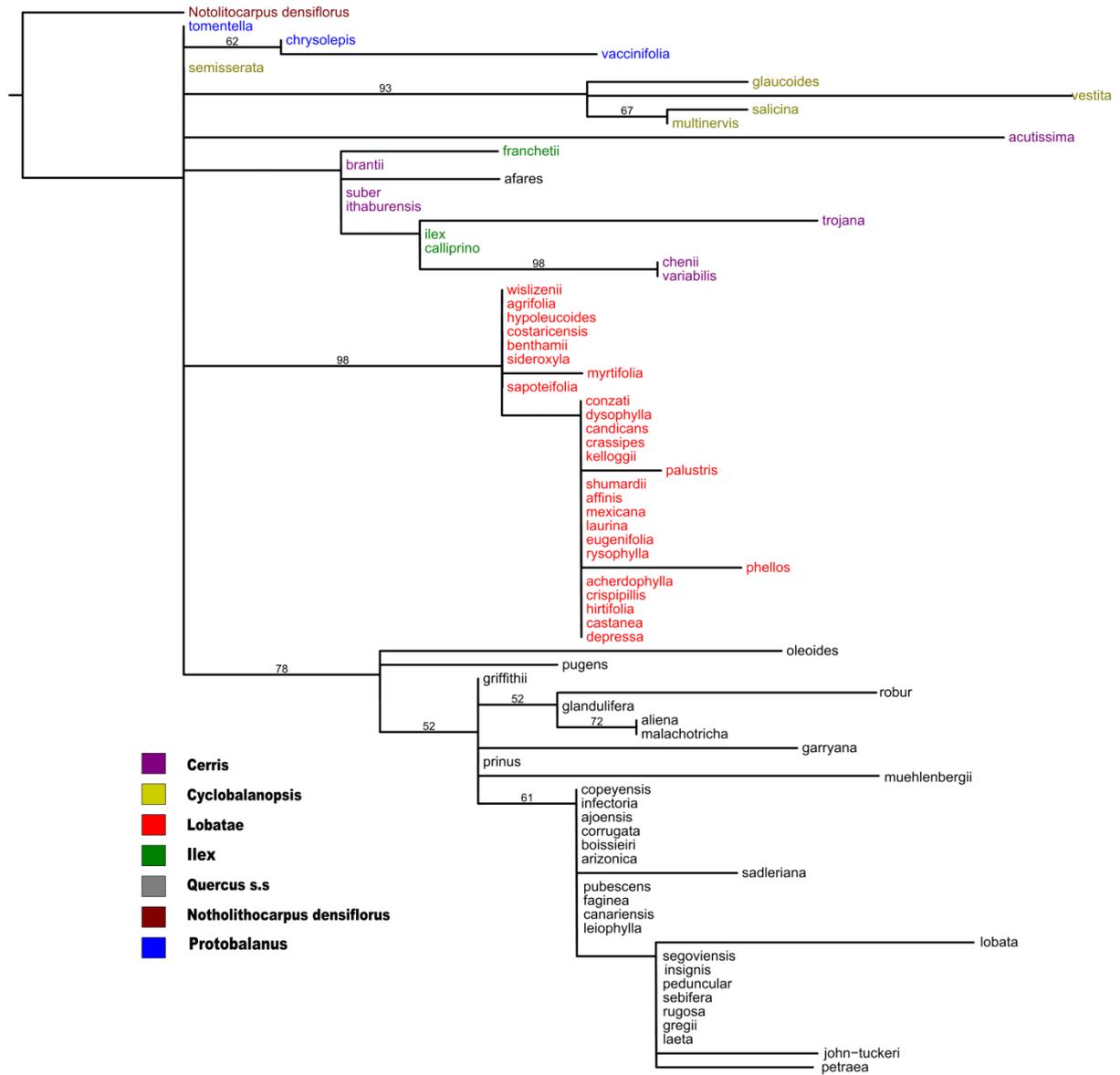
ANNEXE 4E: MAXIMUM LIKELIHOOD PHYLOGRAM OF THE GENUS QUERCUS BASED ON CL8745 AND ROOTED ON THE GENUS CASTANEA. NUMBERS CLOSE TO THE NODES REFER TO MAXIMUM LIKELIHOOD BOOTSTRAP SUPPORT.



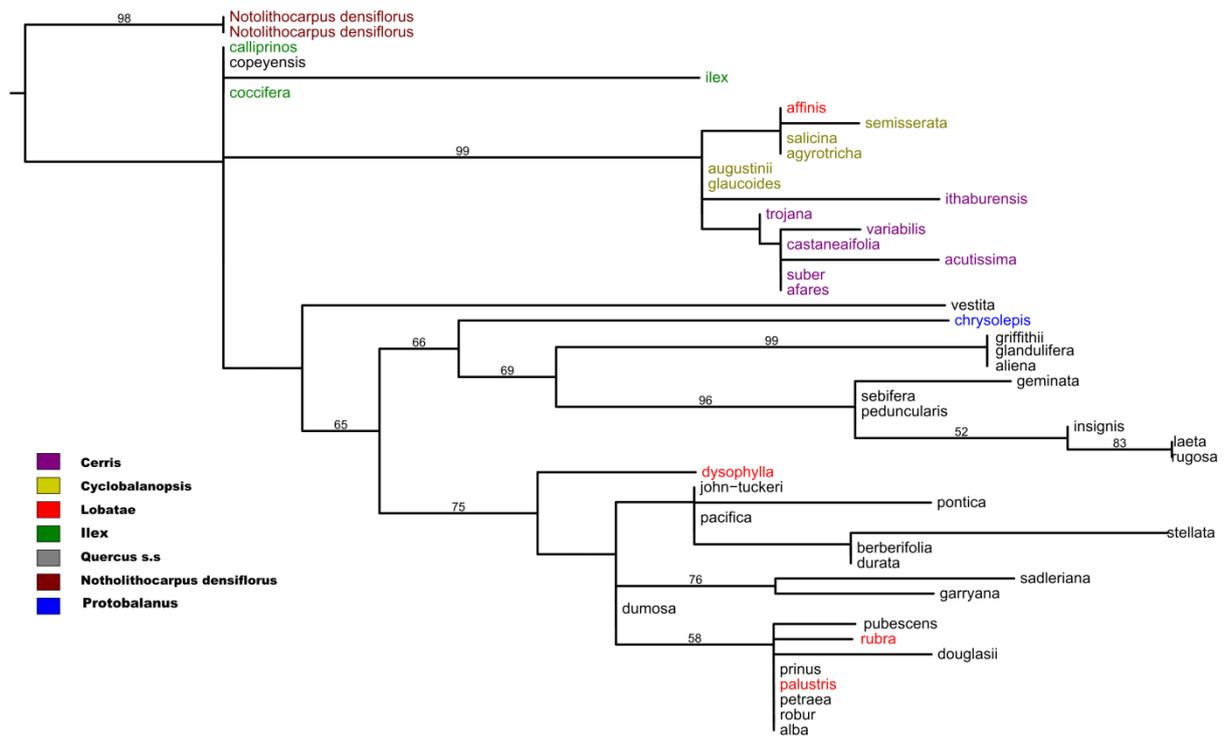
ANNEXE 4F: MAXIMUM LIKELIHOOD PHYLOGRAM OF THE GENUS QUERCUS BASED ON CL9715 AND ROOTED ON THE GENUS CASTANEA. NUMBERS CLOSE TO THE NODES REFER TO MAXIMUM LIKELIHOOD BOOTSTRAP SUPPORT.



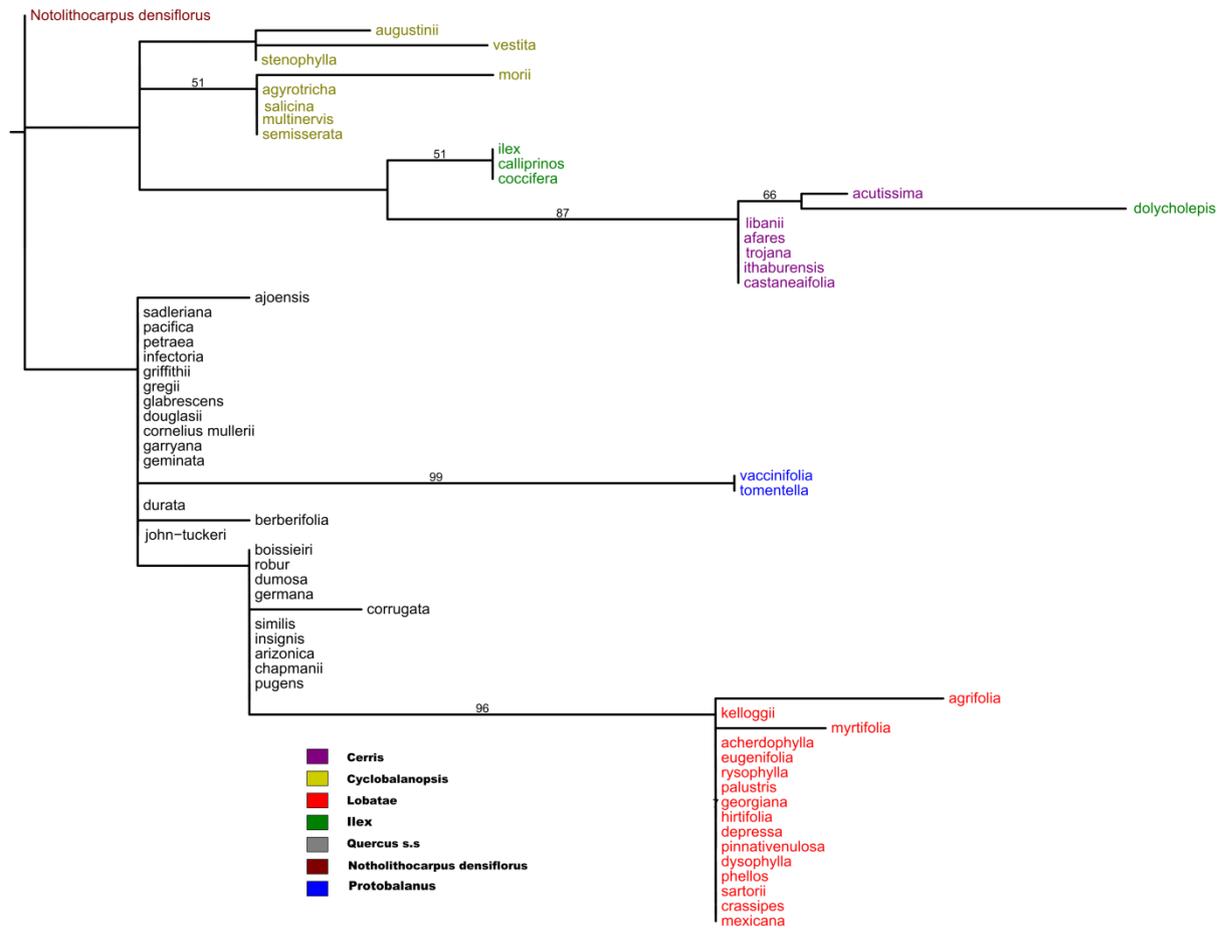
ANNEXE 5A: MAXIMUM LIKELIHOOD PHYLOGRAM OF THE GENUS QUERCUS BASED ON CL1155 AND ROOTED ON THE GENUS CASTANEA. NUMBERS CLOSE TO THE NODES REFER TO MAXIMUM LIKELIHOOD BOOTSTRAP SUPPORT.



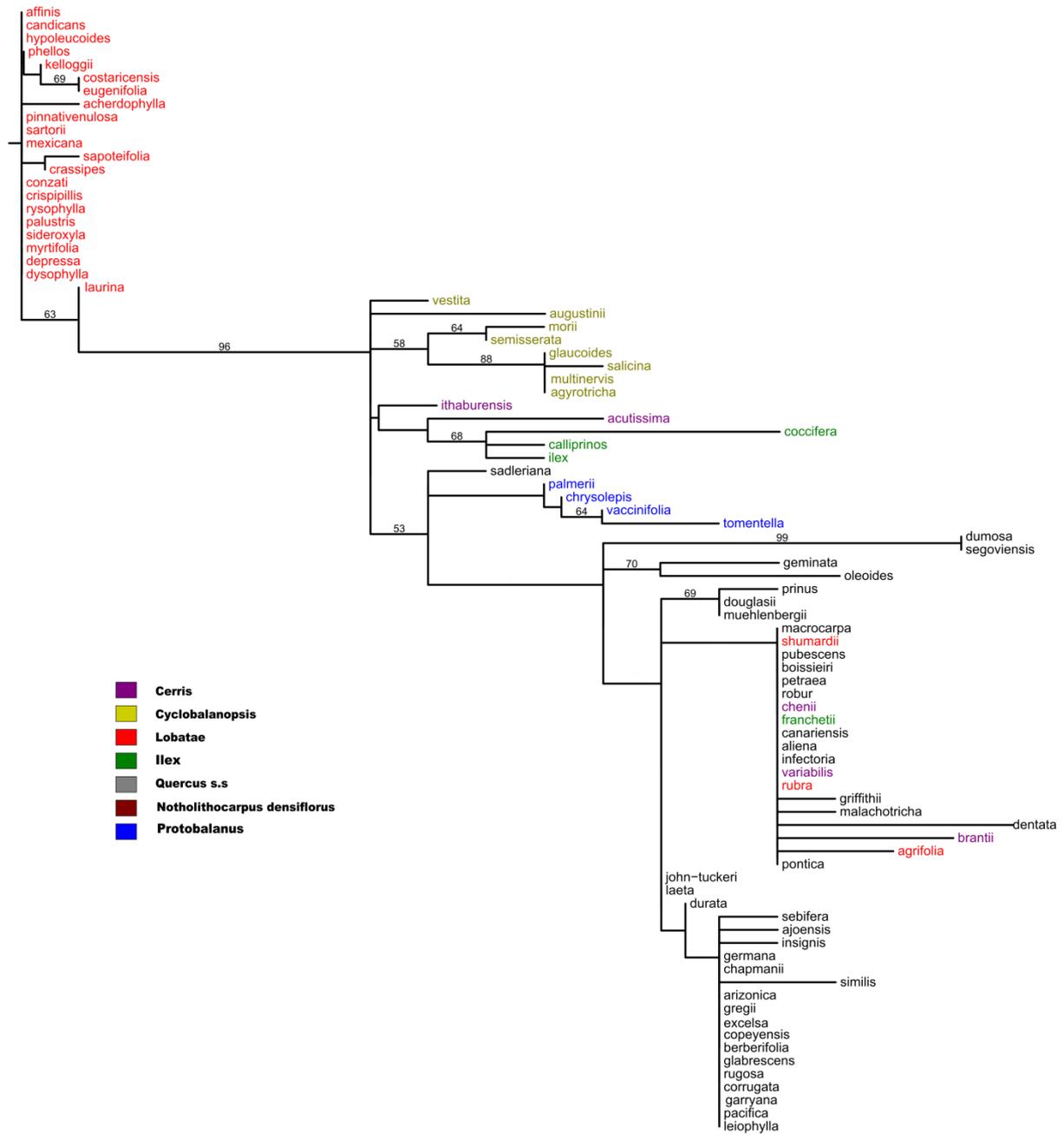
ANNEXE 5B: MAXIMUM LIKELIHOOD PHYLOGRAM OF THE GENUS QUERCUS BASED ON CL8561 AND ROOTED ON THE GENUS CASTANEA. NUMBERS CLOSE TO THE NODES REFER TO MAXIMUM LIKELIHOOD BOOTSTRAP SUPPORT.



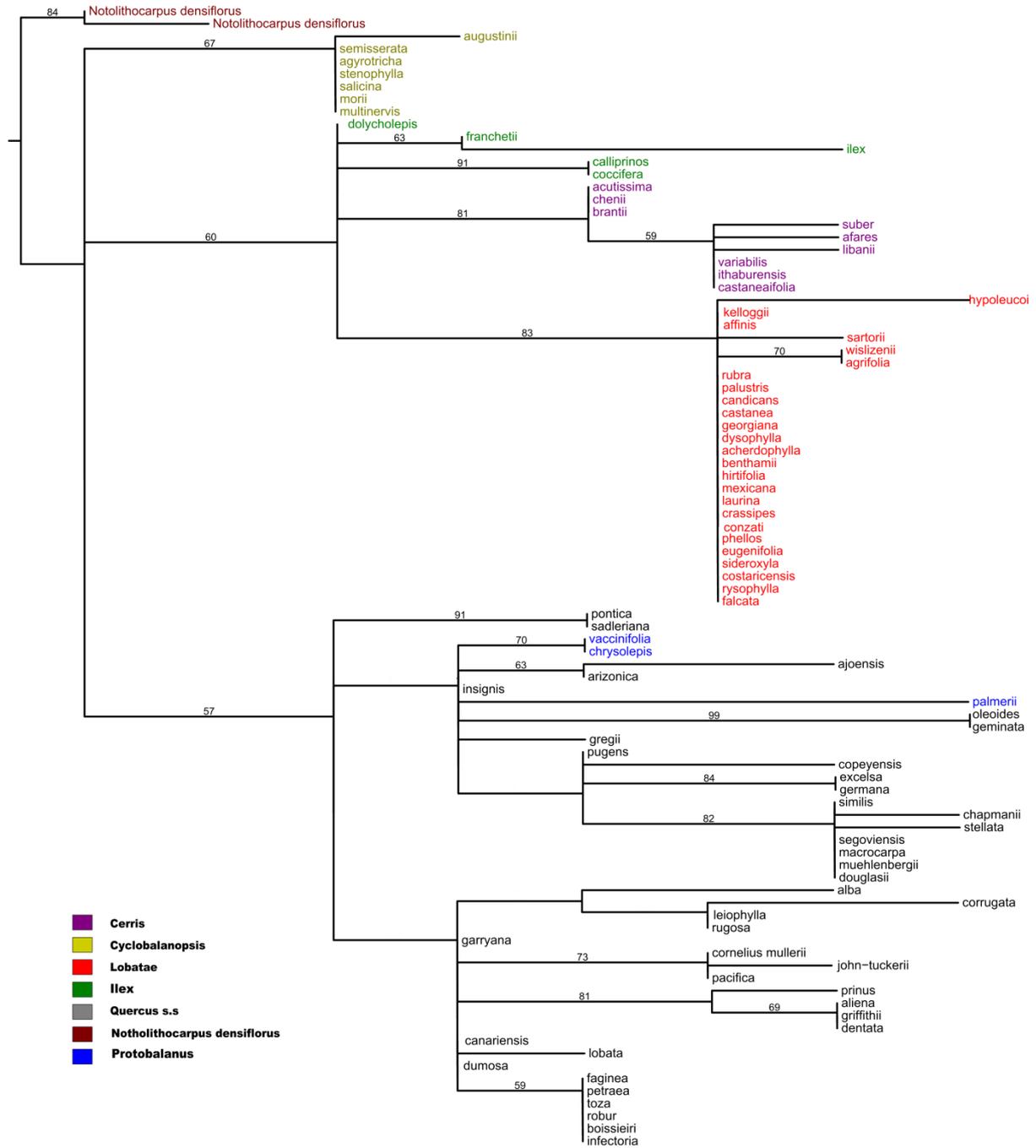
ANNEXE 5C: MAXIMUM LIKELIHOOD PHYLOGRAM OF THE GENUS QUERCUS BASED ON CL5191 AND ROOTED ON THE GENUS CASTANEA. NUMBERS CLOSE TO THE NODES REFER TO MAXIMUM LIKELIHOOD BOOTSTRAP SUPPORT.



ANNEXE 5D: MAXIMUM LIKELIHOOD PHYLOGRAM OF THE GENUS QUERCUS BASED ON CL8461 AND ROOTED ON THE GENUS CASTANEA. NUMBERS CLOSE TO THE NODES REFER TO MAXIMUM LIKELIHOOD BOOTSTRAP SUPPORT.



ANNEXE 5E: MAXIMUM LIKELIHOOD PHYLOGRAM OF THE GENUS QUERCUS BASED ON CL8745 AND ROOTED ON THE GENUS CASTANEA. NUMBERS CLOSE TO THE NODES REFER TO MAXIMUM LIKELIHOOD BOOTSTRAP SUPPORT.



ANNEXE 5F: MAXIMUM LIKELIHOOD PHYLOGRAM OF THE GENUS QUERCUS BASED ON CL9715 AND ROOTED ON THE GENUS CASTANEA. NUMBERS CLOSE TO THE NODES REFER TO MAXIMUM LIKELIHOOD BOOTSTRAP SUPPORT.



## **Reconstructions phylogénétiques du genre *Quercus* à partir de séquences du génome nucléaire et chloroplastique.**

Le genre *Quercus* comprend plus de 500 espèces et est réparti sur l'ensemble de l'hémisphère nord. La phylogénie du genre, faite à ce jour à partir d'un nombre très limité de marqueurs nucléaires, n'était pas résolue. Des incertitudes demeuraient au niveau des nœuds profonds où ont divergé les principaux groupes taxonomiques aujourd'hui reconnus. L'objectif de cette thèse était d'explorer de manière plus exhaustive les ressources génomiques nucléaires et chloroplastiques pour affiner la phylogénie du genre. Les travaux sont basés sur les séquences de six gènes nucléaires et de l'ensemble du génome chloroplastique. Ces travaux confirment le caractère diffus du signal phylogénétique et le gain de résolution obtenu par l'adjonction de séquences nouvelles. Ils confirment également la subdivision du genre en six groupes infragénériques (*Cyclobalanopsis*, *Ilex*, *Cerris*, *Lobatae*, *Quercus s.s.* et *Protobalanus*), dont les relations phylogénétiques ont été précisées, même si certaines irrésolutions persistent. La thèse met très clairement en évidence l'empreinte phylogéographique dans le génome chloroplastique au niveau du genre et de sa distribution mondiale. Le signal phylogéographique chloroplastique ajouté à la phylogénie nucléaire permet d'échafauder un scénario biogéographique de diversification du genre. Ce scénario devra être corroboré par des apports d'autres disciplines (paléontologie et géologie historique).

**Mots clés :** inférence phylogénétique, spéciation rapide, irrésolution, *Quercus*, *Cyclobalanopsis*, *Ilex*, *Cerris*, *Protobalanus*, *Quercus s.s.*, *Lobatae*, phylogénie nucléaire, phylogénie chloroplastique, biogéographie historique.

## **Phylogeographic reconstructions of the genus *Quercus* based on nuclear and chloroplast DNA sequences.**

The genus *Quercus* comprises more than 500 species, and is widely distributed across the Northern hemisphere. Phylogenetic reconstructions based on traditional molecular sequences were so far irresolutive at the deeper nodes where the major extant taxonomic groups have diverged. This thesis aims at improving the phylogeny of the genus by exploring the current nuclear and chloroplast genomic resources. The phylogenetic investigations are based on sequences of six nuclear genes and the entire chloroplast genome. The results confirm that the phylogenetic signal is rather diluted and that substantial improvements can be obtained by adding sequences from additional genes. They also confirm that the genus can be subdivided in six infrageneric groups (*Cyclobalanopsis*, *Ilex*, *cerris*, *Lobatae*, *Quercus s.s.* et *Protobalanus*). Phylogenetic relationships among these groups are refined, although not fully clarified. There is a very clear phylogeographic imprint in the chloroplast genome that extends at the macroevolutionary level at the whole genus across its entire distribution. The phylogeographic structure together with the phylogeny at the nuclear level allows to elaborate an historical scenario of the radiation of the genus. Additional elements coming from other disciplines (paleontology, historical geology) are however necessary to confirm this scenario.

**Keywords :** phylogenetic inference, fast speciation, irresolution, *Quercus*, *Cyclobalanopsis*, *Ilex*, *Cerris*, *Protobalanus*, *Quercus s.s.*, *Lobatae*, nuclear phylogeny, chloroplast phylogeny, historical biogeography.